矩阵理论及其应用

蒋正新
施国梁 编著

北京航空學院出版社

内 容 简 介

本书是专为工科研究生和应用数学专业本科生编写的《矩阵理论及其应用》教材。在内容上除包括了常用的矩阵理论及方法外，还列举了它们在数值分析、最优化、控制论、随机过程及微分方程等方面的应用。以便为学习《泛函分析》奠定一定的基础。全书共分六章：线性代数基础、矩阵的因式分解、广义逆矩阵、矩阵分析、特征值的估计、非负矩阵及其应用。每章都附有适量的例题与习题。

本书也适合于工科学校的教师、工程师以及有关的科技工作者自学之用。

矩阵理论及其应用

蒋正新　施国梁　编著

责任编辑　郭继烈

北京航空学院出版社出版
新华书店北京发行所发行　各地新华书店经售
北京航空学院印刷厂印刷

850×1168　1/32　印张：13　字数：338千字
1988年3月第一版　1988年3月第一次印刷　印数：7000册
ISBN 7-81012-030-1/O·003　定价：2.50元

前　　言

本书是根据我过去七年在北航研究生院为工科研究生讲授《矩阵论》的讲稿修改扩充而成的。由于研究生听课人数逐年增多（每届数百人），急需教材使用，便约請施国梁同志共同加工。由于讲稿以工科研究生为对象，所以不太追求数学上的系统与严谨，而注重矩阵的理论与方法在工程实际中的应用。因此，作为通用教材，一定存在着多方面的问题，甚至是错误的。敬請批评与指正。

本书初稿曾蒙北京工业学院王朝瑞、史荣昌副教授的评审，提出了很好的修改意见，并指出了个别错漏之处。在此，谨向二位表示衷心的感谢。

蒋正新

一九八七年五月

符号说明

A^{+}　　A的Moore-Penrose广义逆

A^{T}　　A的转置

A^{H}　　A的共轭转置

A_L^{-1}　　A的左逆

A_R^{-1}　　A的右逆

$(A)_{ij}$　　矩阵A的第（i，j）元素

$\sigma(A)$　　A的谱，A的所有特征值集合

$\rho(A)$　　A的谱半径，$\max\{|\lambda|\,|\,\lambda\in\sigma(A)\}$

$\gamma(A)$　　A的素性指数

$\delta(A)$　　$\max\{|\lambda|\,|\,\lambda\in\sigma(A),\ \lambda\neq\rho(A)\}$

$\mathscr{N}(A)$　　A的零空间，A的核

$\mathscr{R}(A)$　　A的值域，A的列空间

$\|A\|$　　A的范数，

$A\geqslant 0$　　A的每个元素都是非负的

$A>0$　　A的每个元素都是正的

$A\geqslant B$　　A的每个元素都不小于B的相应元素

$\det A$　　A的行列式

$\mathrm{adj}A$　　A的伴随矩阵

$\mathrm{tr}A$　　A的迹，A的主对角元之和

$\mathbf{R}$　　实数的集合，实直线

$\mathbf{R}^{n}$　　n维实坐标向量空间

$\mathbf{R}^{m\times n}$　　$m\times n$阶实矩阵空间

$\mathbf{R}_r^{m\times n}$　　所有秩为r的$m\times n$阶实矩阵集合

$\mathbf{C}$　　复数集，复平面

$\mathbf{C}^{m\times n}$　　$m\times n$阶复矩阵空间

$\mathrm{Re}\lambda$　　复数λ的实部

$\mathrm{Im}\lambda$	复数λ的虚部
E^n	n维欧氏空间
X^c	子空间X的代数补
$X^\perp$	集合X的正交补
$(x\|y)$	量向x与y的內积
$x\perp y$	向量x与y垂直
$X\cup Y$	X与Y的并集
$X\cap Y$	X与Y的交集
$X\oplus Y$	子空间X与Y的直和
$A\oplus B$	矩阵A与B的直和，$\mathrm{diag}(A, B)$
$L(X, Y)$	X到Y的所有线性变换集合
$\dim(X)$	空间X的维数
$\mathrm{rank}(A)$	矩阵A的秩
$\mathrm{null}(A)$	矩阵A或线性变换A的零度
$\mathrm{span}M$	向量集合M张成的线性子空间
$P_{L,M}$	沿着空间M向空间L的垂直投影
$\bigoplus_{i=1}^{k} A_i$	$\mathrm{diag}(A_1, A_2, \cdots, A_k)$
$\min_{x\in S} f(x)$	$f(x)$在S上的最小值
$\max_{x\in S} f(x)$	$f(x)$在S上的最大值
$\inf_{x\in S} f(x)$	$f(x)$在S上的下确界
$\sup_{x\in S} f(x)$	$f(x)$在S上的上确界
$\mathrm{diag}(a_1, a_2, \cdots, a_n)$	$n\times n$阶对角阵
$\longrightarrow$	蕴含着，或必要条件
$\Longleftrightarrow$	互为充分必要条件
$x\longmapsto y$	给定x，有唯一的y与之对应

目　　录

第一章　线性代数基础

我们总是习惯于在通常的欧几里德空间 E^3 中考虑和研究问题，你能说出其中的道理吗？不言而喻，我们置身于这一空间之中，对于在这种空间描述的问题当然易于体察。也就是说，E^3 中有与我们直觉思维相合的欧氏几何，它具体、形象，能启发我们去想象。还有，我们从中学到大学所学到的初等数学和高等数学的理论与方法，都是在 E^3 中建立起来的。一句话，E^3 中的几何对象我们感到生动具体。E^3 中所采用的数学工具我们感到熟悉。所以在遇到问题时，总希望能象在 E^3 中一样得到几何解释，也希望能用得心应手的理论方法加以解决。不过，随着科学技术的发展，我们要研究的对象已远远超出了 E^3 的范围。例如，在系统工程和现代控制理论中，需要研究状态向量的集合；在信息工程中，需要研究各种序列的集合；在微分方程、积分方程和函数逼近理论中，需要研究各类函数的集合；在数值分析和计算中，则需要研究矩阵或变换的集合；等等。这时，我们自然会提出这样的一个问题：对于一般性的抽象集合，能否也建立起类似于 E^3 中的一些结构，把 E^3 中的某些经典理论和方法也平行地转移过来，从而使来自不同领域的各种复杂问题能有一个统一的数学描述，能有一个简单直观的几何解释，能用与 E^3 中的共同原理加以解决。事实表明，这种想法在一定程度上是可行的。这也是线性代数和泛函分析的出发点之一。但是，要达到这样一个目的，首先要认真地考察 E^3 的数学结构。

从数学上不难看出，E^3 至少有以下几重结构：

集合论方面的结构。 E^3 中的点可以看作三维实坐标向量的集合

$$V=\{x=(\xi_1, \xi_2, \xi_3) \mid \xi_i \in \mathbf{R}\}$$

我们把它叫做空间 E^3 的基集。对 V 的子集引入集合论中的交、并、补（$\cap$、$\cup$、c）三种运算，便成为一个集合代数（Boole 代数）。

拓扑方面的结构。对 V 中的两点

$$x=(\xi_1, \xi_2, \xi_3), \quad y=(\eta_1, \eta_2, \eta_3)$$

按照等式

$$d(x, y)=[(\xi_1-\eta_1)^2+(\xi_2-\eta_2)^2 +(\xi_3-\eta_3)^2]^{1/2}$$

规定距离，也就是在基集 V 上定义距离函数，便成为一个距离空间（这是一个最常用的拓扑空间）。由距离在 V 中定义开集、闭集、邻域、点列的收敛性、映射的连续性等拓扑概念，从而为建立 E^3 上的微分学奠定基础。

代数方面的结构。对 V 中的向量规定加法及与实数的乘法运算：

$$x+y \doteq (\xi_1+\eta_1, \xi_2+\eta_2, \xi_3+\eta_3)$$

$$\lambda \cdot y \doteq (\lambda\xi_1, \lambda\xi_2, \lambda\xi_3)$$

很显然，这两种运算的结果都仍是 V 中的向量，且满足一定的法则。这样便成为实数域 $\mathbf{R}$ 上的一个线性空间。据上述两种运算定义向量的相关性、子空间、基底与维数，以及基变换和坐标变换等。

欧氏几何方面的结构。在 V 中规定向量 x 和 y 的点积：

$$(x|y) \doteq \xi_1\eta_1+\xi_2\eta_2+\xi_3\eta_3$$

再由 $(x|y)=0$ 规定 x 与 y 正交（即垂直）。还可进一步规定向量的长度，向量间的夹角，正交变换，研究 E^3 中的几何对象在正交变换群之下的不变量，即所谓的欧氏几何学。

测度论方面的结构。把 V 中的向量视为 E^3 中的点，对于 E^3 中直线上、平面上、空间中的点集，分别考虑它们所构成的“长

度”、“面积”、“体积”，从而引入抽象测度的概念。这为E^3上的积分学奠定了基础。

上述几种结构在E^3中不是孤立的，而是有着密切的联系。用数学的术语来讲，就是所赋予基集V的几种结构彼此是相容的。我们也看到空间中的某些几何观念是由拓扑结构提供的（如邻域、开集、闭集等）；有一些几何观念是由代数结构提供的（如基底，坐标，维数，基变换与坐标变换等）；还有一些几何观念，如正交性，正交变换等，是由几种结构的结合所提供的；特别是E^3上的微积分，更是几种结构的高度综合。

从以上的分析看，要想把E^3的结构推广到一般的集合，则先要考虑把几种结构分别作相应的推广。本章就按这种想法，先赋予集合以代数结构，使之成为一个线性空间。然后，进一步赋予内积（即E^3中点积的推广），并由之导出几何结构及拓扑结构，使之成为抽象的欧氏空间或酉空间。最后考虑空间上的线性变换。这样便简明扼要地给出线性代数的基础知识。学过这部分内容的读者，作为复习浏览一下就行了。而只学过线性代数初步的读者，要花相当的时间认真阅读，最好能参考更详尽的教程（如北大编《高等代数》或张远达编《线性代数原理》）。

第一节　綫性空間

一、定义及例子

定义1.1.1（线性空间）　一个线性空间是由下面几个部分组成的：

1．一个域F，其元素叫做标量；

2．一个集合V，其元素叫做向量；

3．一个称作为向量加法的运算，记之为“+”。也就是说，对V中的每一对向量x，y，都有V中的一个向量$x+y$与之对应，并叫做x与y的和，且满足

(a) 加法交换律，$x+y=y+x$；

(b) 加法结合律，$x+(y+z)=(x+y)+z$；

(c) 在V中存在唯一的向量0，叫做零向量，使得对一切V中的向量x有$x+0=x$；

(d) 对V中的每个向量x，都存在唯一的V中的向量$-x$，使得$x+(-x)=0$；

4. 一个称作为向量的标量乘法的运算。也就是说，对域F中的每个标量λ和V中的每个向量x，都有V中的一个向量λx与之对应，并叫做λ与x的乘积，且满足：

(e) 在F中存在标量1使得对V中的每个向量x，都有$1x=x$；

(f) 对F中的一切λ，μ和V中的任意x有$(\lambda\mu)x=\lambda(\mu x)$；

(g) 对F中的一切λ和V中的一切x，y有$\lambda(x+y)=\lambda x+\lambda y$；

(h) 对F中的一切λ，μ和V中的一切x有$(\lambda+\mu)x=\lambda x+\mu x$。

线性空间的定义表明：它是域F和向量集合V的复合；在V中定义了向量的加法，在F与V之间定义了标量与向量的乘法，并且两种运算的结果仍落在V中，即V对这两种运算保持封闭；同时这两种运算还要满足八条法则。如果我们把乘法用符号"·"表示，则上述内容可以简记为（V，F，+，·，8），这时称V是域F上的一个线性空间（或向量空间）。V又叫做空间的基集。

由定义可知，若$\{x_1, x_2, \cdots, x_n\}$是线性空间（$V$，$F$，+，·，8）中的任一组向量，$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$是任一组标量，则

$$y \triangleq \lambda_1 x_1+\lambda_2 x_2+\cdots+\lambda_n x_n=\sum_{i=1}^{n}\lambda_i x_i$$

也是V中的向量。这时，我们把y叫做向量组$\{x_1, x_2, \cdots, x_n\}$

的一个**线性组合**，而把 $\sum_{i=1}^{n} \lambda_i x_i$ 叫做 y 的一个**线性表出**。

以后我们只讨论域 F 取实数域 **R** 或复数域 **C** 的情况，但在不强调数域是实的或复的时，仍记为 F。并把线性空间$(V, F, +, \cdot, 8)$ 简记为 V。

例1. 取基集 $V=\{x \mid x=(\xi_1, \xi_2, \cdots, \xi_n)^T, \xi_i \in \mathbf{C}\}$，取标量域 $F=\mathbf{C}$。又设 $y=(\eta_1, \eta_2, \cdots, \eta_n)^T \in V$，$\lambda \in \mathbf{C}$。规定：

1. $x=y \Longleftrightarrow \xi_i=\eta_i$，$i=1, 2, \cdots, n$，
2. $x+y=(\xi_1+\eta_1, \xi_2+\eta_2, \cdots, \xi_n+\eta_n)^T$，
3. $\lambda \cdot x=(\lambda \cdot \xi_1, \lambda \cdot \xi_2, \cdots, \lambda \cdot \xi_n)^T$，
4. $0=(0, 0, \cdots, 0)^T$。

则很容易看出 V 对上述运算保持封闭，并满足八条运算法则。因此 $(V, \mathbf{C}, +, \cdot, 8)$ 是一个复数域上的线性空间，叫做 **n 维复坐标向量空间**，记作 $\mathbf{C}^n$。类似地，将例 1 中的 **C** 全换成 **R**，则可得 n 维实坐标向量空间，记之为 $\mathbf{R}^n$。$\mathbf{C}^n$ 与 $\mathbf{R}^n$ 是最常用的两个线性空间。

值得注意的是，基集 V 仍取例 1 中的 V，而域 F 取实数域 **R**，按例 1 中定义的运算，则 $(V, \mathbf{R}, +, \cdot, 8)$ 是实数域 **R** 上的一个线性空间，它与 $\mathbf{C}^n$，$\mathbf{R}^n$ 是不同的。而反过来，若基集取，

$$V=\{x \mid x=(\xi_1, \xi_2, \cdots, \xi_n)^T, \xi_i \in \mathbf{R}\}$$

而域 F 取复数域 **C**，仍按例 1 规定运算，则不构成一个线性空间。为什么？

例2. 取基集 $V=\{A \mid A=(a_{ij})_{m \times n}, a_{ij} \in \mathbf{C}\}$，取标量域 $F=\mathbf{C}$。又设 $B=(b_{ij}) \in V$，$\lambda \in \mathbf{C}$。规定

1. $A=B \Longleftrightarrow a_{ij}=b_{ij}$，$1 \leqslant i \leqslant m$，$1 \leqslant j \leqslant n$，
2. $A+B=(a_{ij}+b_{ij})$，

3. $\lambda A=(\lambda a_{ij})$,

4. V中的零向量为$m\times n$阶零矩阵。

则容易验证(V, **C**, +, ·, 8)是一个复数域上的线性空间,记之为$\mathbf{C}^{m\times n}$。类似地可定义实数域上的线性空间$\mathbf{R}^{m\times n}$。

例3. 取基集$V=\left\{p(x)\mid p(x)=\sum_{i=0}^{n}\alpha_i x^{n-i},\ \alpha_i\in\mathbf{R}\right\}$,取域$F=\mathbf{R}$,又设$q(x)=\sum_{i=0}^{n}\beta_i x^{n-i}$,$\lambda\in\mathbf{R}$。规定

1. $p(x)=q(x)\Longleftrightarrow\alpha_i=\beta_i$, $i=0, 1, \cdots, n$,

2. $p(x)+q(x)=\sum_{i=0}^{n}(\alpha_i+\beta_i)x^{n-i}$,

3. $\lambda p(x)=\sum_{i=0}^{n}(\lambda\alpha_i)x^{n-i}$,

4. V中的零向量为零多项式。

则不难验证V是**R**上的一个线性空间(V, **R**, +, ·, 8),并记之为$R[x]_n$。如果把V改为

$$V=\left\{p(x)\mid p(x)=\sum_{i=0}^{n}\alpha_i x^{n-i},\ \alpha_i\in\mathbf{R},\ \alpha_0\neq 0\right\}$$

而其它规定不变,则不再构成一个线性空间。为什么?

例4. 对固定的$A\in\mathbf{C}^{m\times n}$,取基集

$$V=\{x\in\mathbf{C}^n\mid Ax=0,\ 0\in\mathbf{C}^m\}$$

取域$F=\mathbf{C}$,并定义与$\mathbf{C}^n$中相同的运算。则(V, **C**, +, ·, 8,)是一个复数域上的线性空间,叫做矩阵A的零空间(或核),也叫做方程$Ax=0$的解空间,记之为$\mathscr{N}(A)$。

例5. 对固定的$A\in\mathbf{C}^{m\times n}$,取基集

$$V=\{y\in\mathbf{C}^m\mid y=Ax,\ x\in\mathbf{C}^n\}$$

取域$F=\mathbf{C}$,并定义与$\mathbf{C}^m$中相同的运算。则(V, **C**, +, ·, 8)是一个复数域上的线性空间,叫做A的列空间,或A的值域,记之为$\mathscr{R}(A)$。

二、基底与维数

定义1.1.2（线性相关） 设V是一个线性空间，$\{x_1, x_2, \cdots, x_n\}$是$V$中的一组向量。如果在$F$中存在一组不全为零的标量$\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$，使得

$$\sum_{i=1}^{n} \alpha_i x_i = 0$$

则称向量组是线性相关的。否则，便说成是线性无关的。由单个零向量构成的向量组$\{0\}$认为是相关的，而单个非零向量认为是线性无关的。

根据定义1.1.2很容易推出：

定理1.1.3 若向量组$\{x_1, x_2, \cdots, x_n\}$是线性相关的，则其中至少有一个向量$x_j (1 \leqslant j \leqslant n)$能用其余的向量线性表出；若该向量组是线性无关的，则其中沒有任何向量能用其余向量线性表出。

上述定理的逆也是成立的，所以也把它当作线性相关和线性无关的定义。

定义1.1.4（空间的基底） 设$\{x_1, x_2, \cdots, x_n\}$是线性空间$V$中的一组线性无关的向量，并且$V$中的每个向量$x$都可用之线性表出，即在$F$中存在一组标量$\{\xi_1, \xi_2, \cdots, \xi_n\}$使得

$$x = \xi_1 x_1 + \xi_2 x_2 + \cdots + \xi_n x_n = \sum_{i=1}^{n} \xi_i x_i$$

则称$\{x_1, x_2, \cdots, x_n\}$为$V$的一个基底，而把标量组$(\xi_1, \xi_2, \cdots, \xi_n)^T$叫做$x$关于基底$\{x_1, x_2, \cdots, x_n\}$的坐标。通常把基底排成一个形式上的矩阵（因为它不一定是矩阵）$\mathscr{B} = (x_1 \ x_2 \ \cdots \ x_n)$。而把坐标视为一个$n$维的坐标向量，当$F = \mathbf{C}$时，坐标属于$\mathbf{C}^n$，当$F = \mathbf{R}$时，坐标属于$\mathbf{R}^n$。

如果$\mathscr{B} = (x_1, x_2, \cdots, x_n)$是空间$V$的一个基底，我们把基底向量的所有可能的线性组合记为

$$S=\left\{x \mid x=\sum_{i=1}^{n} \alpha_i x_i, \ \alpha_i \in F\right\}$$

则由线性空间的定义知$S \subseteq V$；而据基底的定义又有$V \subseteq S$。因此有$V=S$。所以，一个线性空间V是由它的一个基底生成的，或线性张成的。一组向量$\{y_1, y_2, \cdots, y_n\}$（不一定线性无关）线性张成的集合通常有如下记法：

$$\begin{aligned} S &= L(y_1, y_2, \cdots, y_n) \\ &= \mathrm{span}(y_1, y_2, \cdots, y_n) \\ &= [y_1, y_2, \cdots, y_n] \\ &= \left\{x \mid x=\sum_{i=1}^{n} \alpha_i y_i, \ \alpha_i \in F\right\} \end{aligned}$$

因此，$V=\mathrm{span}(x_1, x_2, \cdots, x_n)=[x_1, x_2, \cdots, x_n]$。

如果线性空间$V \neq \{0\}$，则它的基底有无穷多个。但是，它所有的基底都含有相同个数（或基数）的向量。为证明这一论断，先证明一个引理。

引理1.1.5 设$\{x_1, x_2, \cdots, x_n\}$是空间$V$的一个基底，$\{y_1, y_2, \cdots, y_k\}$是$V$的一组线性无关的向量。则从$\{x_1, x_2, \cdots, x_n\}$中至少可以找出一个向量$x_{i_0}(1 \leqslant i_0 \leqslant n)$，使得向量组$\{x_{i_0}, y_2, \cdots, y_k\}$是线性无关的。

证明 设结论不真，即任取x_i $(1 \leqslant i \leqslant n)$，向量组$\{x_i, y_2, \cdots, y_k\}$都是线性相关的。也就是说，存在着一组不全为零的数$\{\alpha_{1i}, \alpha_{2i}, \cdots, \alpha_{ki}\}$，使得

$$\alpha_{1i} x_i + \sum_{l=2}^{k} \alpha_{li} y_l = 0$$

首先可以断定$\alpha_{1i} \neq 0$，否则与$\{y_1, y_2, \cdots, y_k\}$是线性无关的条件矛盾。从而有

$$x_i = -\sum_{l=2}^{k} \frac{\alpha_{li}}{\alpha_{1i}} y_l, \quad i=1, 2, \cdots, n$$

另一方面，由于每个y_l都可用$\{x_1, x_2, \cdots, x_n\}$线性表出，

不妨设$y_1=\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n$，再将前面的$x_i$的表达式代入该式中，便有

$$y_1=-\sum_{i=1}^{n}\beta_i\sum_{l=2}^{k}\frac{\alpha_{li}}{\alpha_{1i}}y_l$$

$$=\sum_{l=2}^{k}\left(-\sum_{i=1}^{n}\beta_i\frac{\alpha_{li}}{\alpha_{1i}}\right)y_l$$

这意味着$\{y_1, y_2, \cdots, y_k\}$线性相关。与原设矛盾，故引理得证。

定理1.1.6 线性空间V的任两个基底所含向量的个数相等。

证明 设$\{x_1, x_2, \cdots, x_n\}$和$\{y_1, y_2, \cdots, y_k\}$是$V$的两个基底，且$k>n$。由引理1.1.5知，在$\{x_1, x_2, \cdots, x_n\}$中必有向量，不妨设为$x_1$(否则,可以对向量组$\{x_1,x_2,\cdots,x_n\}$重新赋序)，使得$\{x_1, y_2, \cdots, y_k\}$线性无关。再据引理1.1.5，又可从$\{x_2, x_3, \cdots, x_n\}$中找到向量，不妨设为$x_2$，使$\{x_1, x_2, y_3,\cdots,y_k\}$线性无关。如此做到第$n$步，便得到线性无关组$\{x_1, x_2, \cdots, x_n, y_{n+1}, \cdots, y_k\}$。但是，由于每个$y_i$都能用基底$\{x_1, x_2, \cdots, x_n\}$线性表出，故向量组$\{x_1, x_2, \cdots, x_n, y_{n+1},\cdots,y_k\}$必是线性相关的。从而导出矛盾。这表明$k>n$是不可能的。同理可证明$k<n$也是不可能的。从而只有$k=n$，即定理得证。

定理1.1.6表明，线性空间的基底所含向量个数是刻画该空间特征的一个数字，我们把它定义为线性空间的维数。

定义1.1.7（维数） 如果线性空间V的基底含有n个向量，则称V是n维空间，记为$\dim V=n$。顺便指出，我们这里只是就n为有限数的情况来定义和证明的，因为线性代数只研究有限维的线性空间。

定理1.1.8 n维线性空间V的任意n个线性无关向量都可以作为V的一个基底。

证明 设$\{x_1, x_2, \cdots, x_n\}$是$V$的一个基底，而$\{y_1, y_2, \cdots, y_n\}$是$V$的一个线性无关组。若定理不真，则意味着在$V$中至少有一个向量$x$不能用$\{y_1, y_2, \cdots, y_n\}$线性表出，因此向量组$\{y_1, y_2, \cdots, y_n, x\}$是线性无关的。但据引理1.1.5可证$\{x_1, x_2, \cdots, x_n, x\}$也是线性无关的，这与$\{x_1, x_2, \cdots, x_n\}$是一个基底矛盾。故$\{y_1, y_2, \cdots, y_n\}$必可线性表出$V$中的一切向量，即为$V$的一个基底。

定理1.1.9 n维空间V中的任意$k(k<n)$个线性无关的向量都可以作为V的一个基底的部分向量。

证明 设$\{x_1, x_2, \cdots, x_k\}$是$V$中的一个线性无关组。则据定理1.1.6知，它不会是V的一个基底。从而至少存在一个向量x_{k+1}，使得$\{x_1, x_2, \cdots, x_k, x_{k+1}\}$仍线性无关。若$k+1=n$，则据定理1.1.8知，$\{x_1, x_2, \cdots, x_k, x_{k+1}\}$为$V$的一个基底，定理得证。若$k+1<n$，如前所述，必有向量$x_{k+2}\in V$使得$\{x_1, x_2, \cdots, x_{k+1}, x_{k+2}\}$线性无关，并且在$k+2=n$时为$V$的一个基底。否则，继续上述的步骤，由于$n=\dim V$是有限的，一定存在某正整数$l$，使得$k+l=n$，即$\{x_1, x_2, \cdots, x_{k+l}\}$为$V$的一个基底。从而定理得证。

例1. 令$e_i=(0, 0, \cdots, 0, \overset{i}{1}, 0, \cdots, 0)^T$，则$\{e_1, e_2, \cdots, e_n\}$是$\mathbf{R}^n$及$\mathbf{C}^n$的一个基底，通常把它叫做$\mathbf{R}^n$及$\mathbf{C}^n$的自然基，它正好是单位方阵$I_n$的$n$个列向量。

例2. 令$E_{ij}\in\mathbf{R}^{m\times n}$是这样一个矩阵，它的第$(i, j)$元素为1，其余元素皆为零。则集合

$$\{E_{ij}\}_{i=1, 2, \cdots, m,\ j=1, 2, \cdots, n}$$

构成$\mathbf{R}^{m\times n}$及$\mathbf{C}^{m\times n}$的一个基底。因为该集合含有$m\cdot n$个矩阵，故$\dim(\mathbf{R}^{m\times n})=\dim(\mathbf{C}^{m\times n})=m\cdot n$。

例3. 不难证明多项式组$\{1, x, \cdots, x^n\}$是$R[x]_n$中的一个线性无关组（请读者证明一下），且$R[x]_n$中的每个多项式

皆可用之线性表出。故它为$R[x]_n$的一个基底。这表明$R[x]_n$是$n+1$维的线性空间。

例4. $\mathbf{R}^{n\times n}$中的任一个可逆矩阵P，其n个列向量p_1，p_2，…，$p_n\in\mathbf{R}^n$构成$\mathbf{R}^n$的一个基底。因为$\{p_1, p_2, \cdots, p_n\}$是线性无关的，且对于任一$y=(\eta_1, \eta_2, \cdots, \eta_n)^T\in\mathbf{R}^n$，令$P^{-1}y=x=(\xi_1\ \xi_2\ \cdots\ \xi_n)^T$，则有

$$y=Px=(p_1\ p_2\ \cdots\ p_n)(\xi_1\ \xi_2\ \cdots\ \xi_n)^T=\sum_{i=1}^{n}\xi_i p_i$$

这表明$\mathbf{R}^n$中的每个向量y皆可用P的列向量线性表出。这个例子也表明，用$\mathbf{R}^n$或$\mathbf{C}^n$的基底所排成的$n\times n$阶方阵是可逆的。

三、基底变换与坐标变换

下面我们给出线性空间的不同基底间及向量关于不同基底的坐标之间的关系，它对研究空间结构及线性变换都十分有用。

定义1.1.10(基底变换阵) 设$\{x_1, x_2, \cdots, x_n\}$及$\{y_1, y_2, \cdots, y_n\}$是线性空间$V$的两个基底，则每个$y_i$都可用$\{x_1, x_2, \cdots, x_n\}$线性表出：

$$y_i=\gamma_{1i}x_1+\gamma_{2i}x_2+\cdots+\gamma_{ni}x_n,\quad i=1, 2, \cdots, n$$

用矩阵记法，则有

$$(y_1\ y_2\ \cdots\ y_n)=(x_1\ x_2\ \cdots\ x_n)P$$

其中$P=(\gamma_{ij})\in F^{n\times n}$，$F^{n\times n}$表示元素取自域$F$的$n\times n$阶矩阵空间。我们把$P$叫做由基底$\{x_1, x_2, \cdots, x_n\}$过渡到基底$\{y_1, y_2, \cdots, y_n\}$的基底变换阵。

很显然，P是可逆的。否则，据线性方程组的理论知，存在非零的n维坐标向量$x=(\xi_1, \xi_2, \cdots, \xi_n\}^T\in F^n$，使得$Px=0$从而有

$$(y_1\ y_2\cdots y_n)x=(x_1\ x_2\cdots x_n)Px=0$$

即$\sum_{i=1}^{n}\xi_i y_i=0$，这与$\{y_1, y_2, \cdots, y_n\}$是线性无关的相矛盾的。矩阵$P^{-1}$是由基底$\{y_1, y_2, \cdots, y_n\}$过渡到基底$\{x_1, x_2, \cdots, x_n\}$的变换阵。

给定向量$x\in V$，它关于基底$\{x_1, x_2, \cdots, x_n\}$和$\{y_1, y_2, \cdots, y_n\}$的唯一坐标分别是$a=(\xi_1, \xi_2, \cdots, \xi_n)^T$和$b=(\eta_1, \eta_2, \cdots, \eta_n)^T$，即

$$x=\sum_{i=1}^{n}\xi_i x_i=\sum_{i=1}^{n}\eta_i y_i$$

写成矩阵形式，即

$$x=(x_1\ \ x_2\cdots x_n)a=(y_1\ \ y_2\cdots y_n)b$$

将$(y_1\ \ y_2\cdots y_n)=(x_1\ \ x_2\cdots x_n)P$代入上式，便有

$$x=(x_1 x_2\cdots x_n)a=(x_1 x_2\cdots x_n)Pb$$

考虑到x关于基底$(x_1 x_2\cdots x_n)$的坐标是唯一的，便有

$$a=Pb \quad 或 \quad b=P^{-1}a$$

这就是向量x关于不同基底下的坐标之间的变换公式。它表明，一旦基底之间的变换阵给定了，坐标之间的变换也就随之确定了。

四、子空间和维数定理

定义1.1.11（子空间） 设V是域F上的线性空间，W是V的一个非空子集。若W中的向量关于V中的加法运算和标量乘法运算也构成域F上的一个线性空间，则称W是V的一个子空间。

定理1.1.12 设V是域F上的线性空间，W是V的一个非空子集。则W是V的一个子空间的充分必要条件为：对任意的x，$y\in W$及任意的$\lambda\in F$，总有$\lambda x+y\in W$。

这个定理的证明是很显然的，留给读者。它表明，要验证非空子集W为V的子空间，只要验证W对两种运算（以后叫做空间V中的线性运算）保持封闭就够了。

按定义，V 及 $\{0\}$ 都是 V 的子空间，叫做平凡子空间。下面举几个常见子空间的例子。

例1. 给定 $A\in \mathbf{R}^{m\times n}$，集合

$$\mathcal{N}(A)=\{x\in \mathbf{R}^n \mid Ax=0\},$$
$$\mathcal{R}(A)=\{y\in \mathbf{R}^m \mid y=Ax,\ x\in \mathbf{R}^n\}$$

分别是 $\mathbf{R}^n$ 及 $\mathbf{R}^m$ 的线性子空间。

例2. 给定线性空间 V 中的一组向量 $\{x_1, x_2, \cdots, x_k\}$，则集合 $\mathrm{span}(x_1, x_2, \cdots, x_k)=\left\{x \mid x=\sum_{i=1}^{k}\lambda_i x_i,\ \lambda_i\in F\right\}$ 是 V 的一个线性子空间，叫做由 $\{x_1, x_2, \cdots, x_k\}$ 线性张成的子空间。

例3. 在通常的三维坐标空间中，过坐标原点的任意直线和平面都是它的一个子空间。读者用矢量的集合给出几个例子，并说明不过原点的直线与平面都不是子空间。

例4. 取 $\mathbf{R}^{n\times n}$ 中的子集

$$S=\{A\in \mathbf{R}^{n\times n} \mid A^T=A\},$$
$$K=\{A\in \mathbf{R}^{n\times n} \mid A^T=-A\}$$

它们都是 $\mathbf{R}^{n\times n}$ 的子空间。值得注意的是，$\mathbf{C}^{n\times n}$ 中的子集

$$S_1=\{A\in \mathbf{C}^{n\times n} \mid A^H=A\},$$
$$K_1=\{A\in \mathbf{C}^{n\times n} \mid A^H=-A\}$$

都不是 $\mathbf{C}^{n\times n}$ 的子空间。为什么？但集合 S_1 与 K_1 却能构成域 $\mathbf{R}$ 上的线性空间，它们既不是 $\mathbf{R}^{n\times n}$ 的子空间，也不是 $\mathbf{C}^{n\times n}$ 的子空间。为什么？

从例2、3可以看出，在一个 $n(>1)$ 维空间中，除了本身及 $\{0\}$ 两个平凡子空间外，它有从1到 $n-1$ 维的子空间存在，且有无穷多。下面再给出几个重要的生成子空间的方式，并指出它们之间的关系。

定义1.1.13 设 X，Y 是线性空间 V 的两个线性子空间，定义集合

$$S=\{x\in V\mid x\in X \text{ 且 } x\in Y\}\triangleq X\cap Y$$

$$U=\{z\in V\mid z=x+y,\ x\in X,\ y\in Y\}\triangleq X+Y$$

定理1.1.14（交空间与和空间） 如定义1.1.13所设，则S及U都是V的线性子空间，分别叫做X和Y的交空间及和空间。

定理的证明很简单，从略。值得注意的是，两个子空间的并集$X\cup Y$，一般不是子空间。它们有如下明显的关系。

定理1.1.15 设X，Y为空间V的两个子空间，则有

1. $X\cap Y \begin{matrix}\subseteq & X \\ \subseteq & Y\end{matrix} \subseteq X+Y\subseteq V$

2. $X+Y=X\Longleftrightarrow Y\subseteq X$

3. $X\cap Y=X\Longleftrightarrow Y\subseteq Y$

4. $X+Y=X\cup Y\Longleftrightarrow a)\ X\subset Y$ 或 $b)\ Y\subset X$

定义1.1.16（交空间与和空间） 设S_1，S_2，…，S_n是空间V的一组子空间，则可归纳地定义交空间与和空间如下：

$$T_n=(\cdots((S_1\cap S_2)\cap S_3)\cdots)\cap S_n$$
$$Q_n=(\cdots((S_1+S_2)+S_3)\cdots)+S_n$$

由于子空间的交运算及和运算具有交换性及结合性，故上述定义又可记成

$$T_n=\bigcap_{i=1}^{n}S_i,\quad Q_n=\sum_{i=1}^{n}S_i$$

定义1.1.17（直接和） 设S_1，S_2是V的子空间，并且$S_1\cap S_2=\{0\}$，则称和空间S_1+S_2为S_1与S_2的直接和，简称直和。记之为$S_1\dot{+}S_2$。

定义1.1.18（直和分解） 如果V的子空间S，S_1、S_2有关系$S=S_1\dot{+}S_2$，则称S有一个直和分解。特别当$V=S_1\dot{+}S_2$时，称S_1和S_2是一对（代数上）互补的子空间，或称S_2是S_1的代数补，反之亦然。

定理1.1.19（维数定理） 设V是域F上的线性空间，S_1，S_2是V的子空间。则有。

$$\dim(S_1+S_2)=\dim S_1+\dim S_2-\dim(S_1\cap S_2)$$

证明 设$\dim(S_1\cap S_2)=k$，取$\{x_1, x_2, \cdots, x_k\}$为$S_1\cap S_2$的一个基底。因为$S_1\cap S_2\subseteq S_i$，$i=1, 2$，故$\{x_1, x_2, \cdots, x_k\}$为$S_1$和$S_2$中的线性无关组。据定理1.1.9，可把$\{x_1, x_2, \cdots, x_k\}$加以扩充得到$S_1$的基底

$$\{x_1, x_2, \cdots, x_k, y_{k+1}, \cdots, y_m\}$$

及S_2的基底

$$\{x_1, x_2, \cdots, x_k, z_{k+1}, \cdots, z_l\}$$

其中$m=\dim S_1$，$l=\dim S_2$很显然，S_1+S_2中的每个向量都能用

$$\{x_1, x_2, \cdots, x_k; y_{k+1}, \cdots, y_m; z_{k+1}, \cdots, z_l\}$$

线性表出。倘能证明该向量组是线性无关的，则说明它是S_1+S_2的一个基底。从而也就证明了

$$\begin{aligned}\dim(S_1+S_2)&=k+(m-k)+(l-k)\\&=m+l-k\\&=\dim S_1+\dim S_2-\dim(S_1\cap S_2)\end{aligned}$$

为此，设有F中的标量

$$\lambda_1, \lambda_2, \cdots, \lambda_k; \mu_{k+1}, \mu_{k+2}, \cdots, \mu_m;$$
$$\sigma_{k+1}, \sigma_{k+2}, \cdots, \sigma_l,$$

使得

$$\sum_{i=1}^{k}\lambda_i x_i+\sum_{j=k+1}^{m}\mu_j y_j+\sum_{r=k+1}^{l}\sigma_r z_r=0 \qquad (*)$$

即有

$$\sum_{i=1}^{k}\lambda_i x_i+\sum_{j=k+1}^{m}\mu_j y_j=-\sum_{r=k+1}^{l}\sigma_r z_r$$

等号左端是S_1中的向量，右端是S_2中的向量，故有

$$-\sum_{r=k+1}^{l}\sigma_r z_r \in S_1\cap S_2$$

从而存在标量δ_1，δ_2，…，$\delta_k\in F$，使得

$$-\sum_{r=k+1}^{l}\sigma_r z_r=\sum_{i=1}^{k}\delta_i x_i$$

由于$\{x_1, x_2, \cdots, x_k, z_{k+1}, \cdots, z_l\}$线性无关，故有$\sigma_r=0$，$r=k+1$，…，$l$；将$\sigma_r=0$代入（＊），便有

$$\sum_{i=1}^{k}\lambda_i x_i+\sum_{j=k+1}^{m}\mu_j y_j=0$$

又由于$\{x_1, x_2, \cdots, x_k, y_{k+1}, \cdots, y_m\}$线性无关，故有

$$\lambda_i=0,\ 1\leqslant i\leqslant k;\ \mu_j=0,\ k+1\leqslant j\leqslant m$$

这便证明了所需要的结果。

推论1.1.20 若S_1+S_2是S_1和S_2的直和，则有

$$\dim(S_1\dot{+}S_2)=\dim S_1+\dim S_2$$

特别是当$V=S_1\dot{+}S_2$时，有

$$\dim V=\dim S_1+\dim S_2$$

前面关于两个子空间的直和定义，可以推广到n个子空间的情形。

定义1.1.21（直和） 设S_1，S_2，…，S_n是空间V的一组子空间，且满足

1. $S=\sum_{i=1}^{n}S_i$，

2. $S_j\cap(\sum_{i\neq j}S_i)=\{0\}$， $1\leqslant j\leqslant n$

则称S是S_1，S_2，…，S_n的直和，记之为

$$S=S_1\dot{+}S_2\dot{+}\cdots\dot{+}S_n\triangleq\sum_{i=1}^{n}{}^{\cdot}S_i$$

特别是当$V=\sum_{i=1}^{n}{}^{\cdot}S_i$时，又称之为$V$的一个直和分解。

值得注意的是，定义中的条件2不能简单地换为 $S_i \cap S_j = \{0\}$，$i \neq j$。关于子空间的直和有如下重要的定理。

定理1.1.22（直和分解） 设S为n维空间V的一个真子空间，则一定存在S的代数补空间M，使得$V = S \dotplus M$。

证明 设 $\dim S = k < n$，并取S的基底$\{x_1, x_2, \cdots, x_k\}$，将其扩充为$V$的一个基底：

$$\{x_1, x_2, \cdots, x_k, y_{k+1}, \cdots, y_n\}$$

令$M = \mathrm{span}(y_{k+1}, \cdots, y_n)$，易知$S \cap M = \{0\}$（請读者给出证明）及$V = S - M$。从而有$V = S \dotplus M$。

下面不加证明地给出两个重要定理，請读者作为课后练习。

定理1.1.23 设S_1与S_2是空间V的子空间，则下面命题是互相等价的。

1. $S_1 \cap S_2 = \{0\}$
2. 对任一$x \in S_1 + S_2$，存在唯一的$x_1 \in S_1$，$x_2 \in S_2$，使得$x = x_1 + x_2$。
3. 若$x_1 \in S_1$，$x_2 \in S_2$且$x_1 + x_2 = 0$，则必有$x_1 = 0$，$x_2 = 0$。
4. 若$\{x_1, x_2, \cdots, x_k\}$，$\{y_1, y_2, \cdots, y_l\}$分别为$S_1$，$S_2$的基底，则$\{x_1, x_2, \cdots, x_k, y_1, \cdots, y_l\}$为$S_1 + S_2$的基底。
5. S_1中的任一线性无关组与S_2中的任一线性无组关的并集为$S_1 + S_2$中的线性无关组。
6. $\dim(S_1 + S_2) = \dim S_1 + \dim S_2$。

定理1.1.24 设S_1，S_2，$\cdots$，S_n为空间V的一组子空间，则下述命题互相等价。

1. $S = \dot{\sum\limits_{i=1}^{n}} S_i$。
2. 对任一$x \in S = \sum\limits_{i=1}^{n} S_i$，它有唯一的表示

$$x=\sum_{i=1}^{n}x_i,\ x_i\in S_i,\ i=1,\ 2,\ \cdots,\ n。$$

3． $S=\sum_{i=1}^{n}S_i$中的零向量有唯一的表示：

$$0=0+0+\cdots+0。$$

4． $\dim\left(\sum_{i=1}^{n}S_i\right)=\sum_{i=1}^{n}\dim S_i$。

5． 取每个 S_i 的一个基底，其并集构成 $\sum_{i=1}^{n}S_i$ 的一个基底。

6． $\left(\sum_{i=1}^{j-1}S_i\right)\cap S_j=\{0\}$，$2\leqslant j\leqslant n$。

第二节　欧氏空間和酉空間

在上节我们用纯代数的方法给出了空间中一些有几何色彩的概念：空间与子空间，基底与维数，坐标与坐标变换等。但是，反映欧氏几何的一些本质的概念，如矢量的长度，矢量之间的夹角，特别是矢量的正交性，标准正交基底，正交变换以及在正交变换群下的不变量等，都还沒有提出。如何在抽象的线性空间中引入上述类似的定义并演绎出相应的结果，就是本节的主要任务。回顾在三维的欧氏空间 E^3 中，也就是在解析几何中，上述的概念都与向量的点积（以后我们改称内积）有直接的联系。譬如，向量$x=(\xi_1,\ \xi_2,\ \xi_3)^T$与$y=(\eta_1,\ \eta_2,\ \eta_3)^T$ 的点积定义为

$$(x|y)=\xi_1\eta_1+\xi_2\eta_2+\xi_3\eta_3=y^Tx,$$

而向量x的长度（以后改称为范数）定义为

$$\|x\|=(\xi_1^2+\xi_2^2+\xi_3^2)^{1/2}=(x|x)^{1/2},$$

向量x与y之间的夹角θ的余弦为

$$\cos\theta = \frac{(x|y)}{\|x\|\,\|y\|}$$

$$= \frac{\xi_1\eta_1 + \xi_2\eta_2 + \xi_3\eta_3}{\sqrt{\xi_1^2+\xi_2^2+\xi_3^2}\ \sqrt{\eta_1^2+\eta_2^2+\eta_3^2}},$$

特别是，当$(x|y)=0$时，规定$x\perp y$。还有点x与y之间的距离定义为

$$d(x,y) = [(\xi_1-\eta_1)^2+(\xi_2-\eta_2)^2+(\xi_3-\eta_3)^2]^{1/2}$$
$$= \|x-y\|$$

这表明，一旦有了点积的定义，其它的几个概念便逐个诱导出来。所以，我们首先想到要在抽象的线性空间中引入向量内积的概念。为此，就需要考查E^3中向量点积的根本性质。很容易看出，两个向量的点积是一个标量（即实数），所以可以认为$(x|y)$是定义在E^3上的一个二元实函数，且满足下述的性质：

1. 正定性：$(x|x)\geqslant 0$，当且仅当$x=0$时$(x|x)=0$；
2. 对称性：$(x|y)=(y|x)$；
3. 可加性：$(x+y|z)=(x|z)+(y|z)$；
4. 齐次性：$(\lambda x|y)=\lambda(x|y)$。

当然，点积还有其它的性质。数学家经过相当长时间的反复研究和归纳，认为这些性质是最根本的，可以把它当作公理体系在抽象的线性空间中定义内积。

一、内积空间和例子

定义1.2.1（内积） 设F是实数域或复数域，V是F上的一个线性空间。所谓V上的一个内积，是指定义在V上的一个二元函数，记之为$(x|y)$，且对一切x，y，$z\in V$及$\lambda\in F$满足：

1. 正定性：$(x|x)\geqslant 0$，当且仅当$x=0$时有$(x|x)=0$；
2. 共轭对称性：$(x|y)=\overline{(y|x)}$，其中$\overline{(y|x)}$表示对复数$(y|x)$取共轭。

3．关于第一个变元的可加性：

$$(x+y|z)=(x|z)+(y|z);$$

4．关于第一个变元的齐次性：

$$(\lambda x|y)=\lambda(x|y)。$$

以上便是内积的公理化的定义。从2，3，4很容易推出下述有用的性质：

5．$\left(\sum_{i=1}^{n}\lambda_i x_i|y\right)=\sum_{i=1}^{n}\lambda_i(x_i|y)$；

6．$\left(x|\sum_{j=1}^{m}\mu_j y_j\right)=\sum_{j=1}^{m}\bar{\mu}_j(x|y_j)$，

7．$\left(\sum_{i=1}^{n}\lambda_i x_i|\sum_{j=1}^{m}\mu_j y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_i\bar{\mu}_j(x_i|y_j)$

性质5叫做内积关于第一个变元是线性函数；性质6叫做内积关于第二个变元是共轭线性的，有时又叫做半个线性。特别当$F=$ **R**时，性质6，7中的$\bar{\mu}_j=\mu_j$，性质2中的共轭对称性变成对称性。这时内积$(x|y)$是一个对称正定的双线性函数（或泛函）。可以这样说：实线性空间上的内积是一个对称正定的双线性泛函；而复线性空间上的内积是$1\frac{1}{2}$线性的泛函。

定义1.2.2（内积空间） 域F上的线性空间V，如果在其上定义了内积，则称之为内积空间，记之为$(V, F, +, \cdot, (\cdot|\cdot))$，或简记为$(V, (\cdot|\cdot))$。

例1. 在$\mathbf{C}^n$上，把向量$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$，$y=(\eta_1, \eta_2, \cdots, \eta_n)^T$的内积定义为

$$(x|y)\triangleq y^H x=\xi_1\bar{\eta}_1+\xi_2\bar{\eta}_2+\cdots+\xi_n\bar{\eta}_n$$

很容易验证它满足内积的四条公理，故$(\mathbf{C}^n, (\cdot|\cdot))$是一个复（数域上的）内积空间。

例2. 在$\mathbf{R}^n$上，把向量$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$，$y=(\eta_1, \eta_2, \cdots, \eta_n)^T$的内积定义为

$$(x|y) \hat{=} y^T x = \xi_1\eta_1 + \xi_2\eta_2 + \cdots + \xi_n\eta_n$$

则（$\mathbf{R}^n$，$(\cdot|\cdot)$）是一个实（数域上的）内积空间。

例3. 在$\mathbf{C}^{m\times n}$上，把矩阵$A=(a_{ij})$，$B=(b_{ij})$的内积定义为

$$(A|B) \hat{=} \mathrm{tr}(B^H A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\bar{b}_{ij}$$

则（$\mathbf{C}^{m\times n}$，$(\cdot|\cdot)$）是一个复内积空间。而在$\mathbf{R}^{m\times n}$上内积定义为

$$(A|B) \hat{=} \mathrm{tr}(B^T A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

则（$\mathbf{R}^{m\times n}$，$(\cdot|\cdot)$）是一个实内积空间。

例4. 在$R[x]_n$上，把多项式$p(x)=\sum_{i=0}^{n}\alpha_i x^{n-i}$，$q(x)=\sum_{i=0}^{n}\beta_i x^{n-i}$的内积定义为。

$$(p|q) \hat{=} \int_0^1 p(x)q(x)dx$$

则容易验证（$R[x]_n$，$(\cdot|\cdot)$）也构成一个实内积空间。

定义1.2.3（欧氏空间与酉空间） 有限维的实内积空间叫做欧氏空间（Euclidean Space）；有限维的复内积空间叫做酉空间（Unitary Space）。

上面例中的（$\mathbf{R}^n$，$(\cdot|\cdot)$），（$\mathbf{R}^{m\times n}$，$(\cdot|\cdot)$），（$R[x]_n$，$(\cdot|\cdot)$）都是欧氏空间。而（$\mathbf{C}^n$，$(\cdot|\cdot)$），（$\mathbf{C}^{m\times n}$，$(\cdot|\cdot)$）是酉空间的例子。在这些例子中，都是按照最常用的方式定义内积的，读者也很容易验证它们分别都满足内积的四条公理。在工程实际中，根据特殊的需要，往往用其它方式定义内积，遇到时我们再指出。

二、与内积有关的概念及关系式

我们引入内积的目的是希望在线性空间中的进一步诱导出几何和拓扑结构，使之更象通常的三维欧氏空间。在本节的引言中我们可以看出，在 E^3 中向量的内积，向量的夹角，向量的长度之间自然的满足一些条件。譬如：

$$|\cos\theta| = \left|\frac{(x|y)}{\|x\|\,\|y\|}\right| \leqslant 1,$$

$$\|x+y\| \leqslant \|x\| + \|y\|,$$

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

它们有明显的几何解释：夹角的余弦应该小于或等于1；三角形的两边之和应该大于第三边；平行四边形对角线的平方和等于四条边的平方之和。当然还有勾股定理等没有写出。所以，要想说明我们抽象定义的内积也能合理地诱导出所要求的结构，就必须证明它满足类似上述的性质。

引理1.2.4 设 $(V, (\cdot|\cdot))$ 是域 F 上的内积空间，并记 $\|x\| = (x|x)^{1/2}$ （因为$(x|x)\geqslant 0$，故$(x|x)^{1/2}$ 有意义）。则对一切x，y，$z\in V$ 及$\lambda\in F$，有

(1) $\|\lambda x\| = |\lambda|\,\|x\|$；

(2) $\|x-y\|^2 + \|x+y\|^2 = 2\|x\|^2 + 2\|y\|^2$；

(3) $|(x|y)| \leqslant \|x\|\,\|y\|$；

(4) $\|x+y\| \leqslant \|x\| + \|y\|$；

证明 (1) 与 (2) 从内积的四条公理直接可以推出。对于(3)，当$x=0$或$y=0$时，显然成立。现假定$x\neq 0$，即$\|x\|\neq 0$令

$$u = y - \frac{(y|x)}{\|x\|^2}x$$

显然有 $(u|x)=0$，故有

$$0 \leqslant \| u \|^2 = (u | u)$$

$$= (y|y) - \frac{(y|x)(x|y)}{\| x \|^2}$$

$$= \| y \|^2 - \frac{|(x|y)|^2}{\| x \|^2}$$

即

$$|(x|y)|^2 \leqslant \| x \|^2 \| y \|^2$$

$$|(x|y)| \leqslant \| x \| \| y \|$$

这便证明了（3）。利用（3），便有

$$\| x+y \|^2 = (x+y|x+y)$$

$$= \| x \|^2 + (x|y) + (y|x) + \| y \|^2$$

$$= \| x \|^2 + 2\mathrm{Re}(x|y) + \| y \|^2$$

$$\leqslant \| x \|^2 + 2\| x \| \| y \| + \| y \|^2$$

$$= (\| x \| + \| y \|)^2$$

这便证明了（4）。从证明过程中还可以看出，不等式$|(x|y)| \leqslant \| x \| \| y \|$当且仅当$y = (y|x)x / \| x \|^2$时，才有等号成立，也就是说当且仅当$x$与$y$线性相关时等号才成立。这个不等式是数学上有名的柯西-许瓦尔兹-布尼雅可夫斯基（Cauchy-Schwayz-Буняковский）不等式，它有极广泛的应用。

在内积空间（$\mathbf{C}^n$，$(\cdot|\cdot)$，（$\mathbf{R}^n$，$(\cdot|\cdot)$），（$\mathbf{C}^{m\times n}$，$(\cdot|\cdot)$），（$\mathbf{R}^{m\times n}$，$(\cdot|\cdot)$），（$R[x]$，$(\cdot|\cdot)$）中，它们的柯西-许瓦尔兹-布尼雅可夫斯基不等式分别为

$$\left|\sum \xi_i \eta_i\right| \leqslant \left(\sum |\xi_i|^2\right)^{1/2} \left(\sum |\eta_i|^2\right)^{1/2}$$

$$|\mathrm{tr}(B^H A)| \leqslant (\mathrm{tr}(A^H A))^{1/2} (\mathrm{tr}(B^H B))^{1/2}$$

$$\left|\int_0^1 p(x)q(x)\,dx\right|$$

$$\leqslant \left(\int_0^1 |p(x)|^2 dx\right)^{1/2} \left(\int_0^1 |q(x)|^2 dx\right)^{1/2}$$

定义1.2.5（正交性） 内积空间（V，$(\cdot|\cdot)$）中的两个向

量 x，y，如果有 $(x|y)=0$，则称它们是互相正交的，记之为 $x\perp y$。很显然，零向量与空间中的每个向量都正交。值得注意的是，只有零向量才与每个向量都正交（根据内积的第一条公理），这个结论后面要用到。特别对欧氏空间中的两个非零矢量 x，y，其夹角 θ 定义为

$$\theta=\arccos\frac{(x|y)}{\|x\|\ \|y\|}$$

据 $C-S-\delta$ 不等式，这个定义是有意义的。

根据正交性的定义，很容易证明：如果 x_1，x_2，…，x_n 是内积空间 $(V,(\cdot|\cdot))$ 中两两互相正交（即 $(x_i|x_j)=0$，$i\neq j$）的一组向量，则有

$$\left\|\sum_{i=1}^{n}x_i\right\|^2=\sum_{i=1}^{n}\|x_i\|^2$$

这就是通常勾股定理的推广。

根据引理1.2.4中的（4），容易证明：对内积空间中的任一组向量 x_1，x_2，…，x_n，都有

$$\left\|\sum_{i=1}^{n}x_i\right\|\leqslant\sum_{i=1}^{n}\|x_i\|$$

这就是推广了的三角不等式。由（4）还可证明常用的不等式：

$$|\,\|x\|-\|y\|\,|\leqslant\|x-y\|,$$
$$\|x-y\|\leqslant\|x-z\|+\|z-y\|$$

三、标准正交基底与Gram-Schmidt过程

在空间解析几何中，我们常取两两正交的单位向量 i，j，k 来线性表出空间的所有向量，这给问题的研究带来极大的方便。试问：在抽象的欧氏空间或酉空间中，能否也找到这样一组向量？它们又具备那些优点？下面就来回答这个问题。

定义1.2.6 在内积空间 $(V,(\cdot|\cdot))$ 中，一组不含零向量的向量组 $\{x_1, x_2, \cdots, x_n\}$，如果两两互相正交，则称之为

一个正交向量组。

定理1.2.7 正交向量组是线性无关的。

证明 设存在标量$\lambda_i \in F$，$i=1, 2, \cdots, n$，使得

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n = 0$$

则对任一$x_i (i=1, 2, \cdots, n)$有

$$(\lambda_1 x_1 + \cdots + \lambda_n x_n | x_i) = (0 | x_i) = 0$$

据内积关于第一变元的线性性知，有

$$\sum_{j=1}^{n} \lambda_j (x_j | x_i) = 0$$

再据正交性，当$i \neq j$时 $(x_j | x_i) = 0$，故有

$$\lambda_i (x_i | x_i) = 0, \quad i=1, 2, \cdots, n,$$

又因 $(x_i | x_i) > 0$，故$\lambda_i = 0$，$i=1, 2, \cdots, n$。

这个定理一方面说明，在n维的内积空间中，两两互相正交的非零向量不会超过n个。另一方面也说明，任意n个两两正交的非零向量都可作为该空间的一个基底。

定义1.2.8 n维内积空间中的n个向量$\{x_1, x_2, \cdots, x_n\}$如果满足

$$(x_i | x_j) = \delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$$

则称之为该空间的一个标准正交基底。其中δ_{ij}叫做 Kronecker *delta*。它表明诸x_i是两两正交且为单位长的向量。

每个有限维的内积空间是否都有标准正交基底呢？下面我们通过 Gram-Schmidt 过程具体构造出来，从而也就回答了这一问题。

定理1.2.9 每个n维内积空间 $(V, (\cdot|\cdot))$ 一定存在标准正交基底。

证明 设$\{x_1, x_2, \cdots, x_n\}$是$V$的一个基底。首先

令 $$y_1 = x_1, \quad z_1 = \frac{y_1}{\|y_1\|},$$

我们下面在子空间span(x_1, x_2)中寻求一个向量y_2，使得$y_2 \perp z_1$。故可令

$$y_2 = x_2 + \lambda_{21} z_1$$

其中λ_{21}可据$y_2 \perp z_1$定出。因为x_2，z_1是线性无关的，故$y_2 \neq 0$。由下式

$$\begin{aligned} 0 &= (y_2 | z_1) = (x_2 + \lambda_{21} z_1 | z_1) \\ &= (x_2 | z_1) + \lambda_{21}(z_1 | z_1) \end{aligned}$$

定出$\lambda_{21} = -(x_2 | z_1)/\|z_1\|^2 = -(x_2 | z_1)$，从而有

$$y_2 = x_2 - (x_2 | z_1) z_1, \quad z_2 \triangleq \frac{y_2}{\|y_2\|}$$

再从子空间span$(x_1, x_2, x_3) = $span$(z_1, z_2, x_3)$中寻求向量$y_3$，使得$y_3 \perp z_1$且$y_3 \perp z_2$。为此，令

$$y_3 = x_3 + \lambda_{31} z_1 + \lambda_{32} z_2$$

显然x_3，z_1，z_2是线性无关的，故$y_3 \neq 0$，由下式

$$0 = (y_3 | z_1) = (x_3 | z_1) + \lambda_{31}(z_1 | z_1)$$

$$0 = (y_3 | z_2) = (x_3 | z_2) + \lambda_{32}(z_2 | z_2)$$

可解出$\lambda_{31} = -(x_3 | z_1)$，$\lambda_{32} = -(x_3 | z_2)$。从而有

$$y_3 = x_3 - \sum_{i=1}^{2} (x_3 | z_i) z_i, \quad z_3 \triangleq \frac{y_3}{\|y_3\|}$$

按照归纳原理，设已经求出两两正交且为单位长的向量组z_1，z_2，…，z_k，并满足

$$\text{span}(x_1, x_2, \cdots, x_k) = \text{span}(z_1, z_2, \cdots, z_k)$$

则在子空间span$(x_1, x_2, \cdots, x_k, x_{k+1}) = $span$(z_1, z_2, \cdots, z_k, x_{k+1})$寻求$y_{k+1}$，使得$y_{k+1} \perp z_i$，$i = 1, 2, \cdots, k$。为此，设

$$y_{k+1} = x_{k+1} + \sum_{i=1}^{k} \lambda_{k+1\,i} z_i$$

同理可从方程组

$$0 = (y_{k+1} | z_i) = (x_{k+1} | z_i) + \lambda_{k+1\,i}$$

$$i = 1, 2, \cdots, k$$

定出$\lambda_{k+1,i}=-(x_{k+1}|z_i)$。$i=1,2,\cdots,k$。从而有

$$y_{k+1}=x_{k+1}-\sum_{i=1}^{k}(x_{k+1}|z_i)z_i,\quad z_{k+1}=\frac{y_{k+1}}{\|y_{k+1}\|}$$

当$k+1=n$时，便得V的一个标准正交基底

$$\{z_1,z_2,\cdots,z_n\}$$

在求$\{z_1,z_2,\cdots,z_n\}$的过程中，我们有两个步骤，首先求正交组$\{y_1,y_2,\cdots,y_n\}$，中间又穿插了求$z_i=y_i/\|y_i\|$的程序。前者叫正交化过程，后者叫标准化过程，整个过程叫做Gram-Schmidt 标准正交化过程。一般教科书上是先正交化，最后标准化，这里是穿插进行，为的后面容易得出矩阵的QR分解式。

从定理1.2.9的证明过程，可以总结出下面两点：

1. 给定内积空间V的一个基底$\{x_1,x_2,\cdots,x_n\}$，通过G-S过程可以得到一个标准正交基底：

$$\{z_1,z_2,\cdots,z_n\}$$

两个基底之间的转换关系为：

$$y_1=x_1,\qquad z_1=y_1/\|y_1\|$$

$$y_k=x_k-\sum_{i=1}^{k-1}(x_k|z_i)z_i,\quad z_k=y_k/\|y_k\|,$$

$$k=2,3,\cdots,n。$$

或者

$$x_1=\|y_1\|z_1,$$

$$x_k=\sum_{i=1}^{k-1}(x_k|z_i)z_i+\|y_k\|z_k,\quad k=2,3,\cdots,n$$

用矩阵的记法便有

$$(x_1x_2\cdots x_n)=(z_1z_2\cdots z_n)\begin{pmatrix} \|y_1\| & (x_2|z_1)\cdots & (x_n|z_1) \\ 0 & \|y_2\|\cdots & (x_n|z_2) \\ 0 & 0\ \ \cdots & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & (x_n|z_{n-1}) \\ 0 & 0\ \ \cdots & \|y_n\| \end{pmatrix}$$

上式右端的上三角阵，由于$\|y_i\|\neq 0$，故是一个满秩阵，记之为R，则它是两个基底之间的变换阵。

2． G-S过程表明：

$$\text{span}(x_1, x_2, \cdots, x_k)=\text{span}(z_1, z_2, \cdots, z_k),$$
$$1\leqslant k\leqslant n,$$

它有明显的几何意义：G-S过程是先在一维子空间$[x_1]$上取单位向量z_1，然后在二维子空间$[x_1, x_2]$上取与z_1正交的单位向量z_2，再在三维子空间$[x_1, x_2, x_3]$上取与z_1，z_2都正交的单位向量z_3，依此类推，最后便求出V的一个标准正交基底。

通过下面简单的比较，可看出在内积空间采用标准正交基底比用一般基底有明显的优越性。

第一，$x\in V$关于一般基底$\{x_1, x_2, \cdots, x_n\}$的表示若为$x=\sum_{i=1}^{n}\xi_i x_i$，则其坐标分量$\xi_i$的计算一般是比较困难的。而$x$关于标准正交基底$\{z_1, z_2, \cdots, z_n\}$的表示若为

$$x=\sum_{i=1}^{n}\eta_i z_i$$

则上式两端分别与z_j取内积，利用内积的线性性及基向量的正交性，便可求得

$$(x|z_j)=\left(\sum_{i=1}^{n}\eta_i z_i\middle|z_j\right)=\sum_{i=1}^{n}\eta_i(z_i|z_j)$$

$$=\sum_{i=1}^{n}\eta_i\delta_{ij}=\eta_j$$

$$j=1, 2, \cdots, n$$

又因$(x|z_i)=\|x\|\|z_i\|\cos\theta=\|x\|\cos\theta$，其中$\theta$是$x$与$z_i$之间的夹角。这表明$x$关于标准正交基底的坐标分量$\eta_i$正好是$x$在$z_i$方向上的垂直投影的长度，与解析几何中的表示一致，即

$$x=\sum_{i=1}^{n}(x|z_i)z_i$$

第二，若向量x，$y\in V$关于一般基底$\{x_1, x_2, \cdots, x_n\}$的表示分别为

$$x=\sum_{i=1}^{n}\xi_i x_i,\quad y=\sum_{j=1}^{n}\eta_j x_j$$

则有

$$(x|y)=\left(\sum_{i=1}^{n}\xi_i x_i \,\middle|\, \sum_{j=1}^{n}\eta_j x_j\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\bar{\eta}_j(x_i|x_j)$$

若记$u=(\xi_1, \xi_2, \cdots, \xi_n)^T$，$v=(\eta_1, \eta_2, \cdots, \eta_n)^T$，

$$A=\begin{pmatrix}(x_1|x_1) & (x_1|x_2) & \cdots & (x_1|x_n)\\(x_2|x_1) & (x_2|x_2) & \cdots & (x_2|x_n)\\ & & \vdots & \\(x_n|x_1) & (x_n|x_2) & \cdots & (x_n|x_n)\end{pmatrix}$$

则有$(x|y)=v^H Au$。根据内积的第二条公理，可知$A^H=A$，它是一个Hermite矩阵。有时又把A记作为$G(x_1, x_2, \cdots, x_n)$，叫做向量组$\{x_1, \cdots, x_n\}$的格拉姆(Gram)矩阵。由于当$x\neq 0$时有$u\neq 0$，故有$0<(x|x)=u^H Au$。这说明内积空间V中的每个基底的格拉姆矩阵都是正定的Hermite矩阵。而空间中每对向量的内积都是一个形如$v^H Au$的二次型函数。但是，若取V的标准正交基底$\{z_1, z_2, \cdots, z_n\}$，由于$(z_i|z_j)=\delta_{ij}$，则它的格拉姆矩阵$G(z_1, z_2, \cdots, z_n)=I$。如果$x$，$y$关于该基底的坐标仍记为$u$和$v$，则有$(x|y)=v^H u=\sum_{i=1}^{n}\xi_i\bar{\eta}_i$。若$V$是

欧氏空间，则更有 $(x|y)=\sum_{i=1}^{n}\xi_i\eta_i$。这表明两个向量的内积等于它们的坐标分量对应相乘再相加，与解析几何中向量点积的定义完全相符。

当然，还有其它方面的优越性，比如关于正交变换和求解最小二乘问题，以后遇到时就会明白。

例1. 在具有标准内积的空间 $(\mathbf{R}^n, (\cdot|\cdot))$ 及 $(\mathbf{C}^n, (\cdot|\cdot))$ 中，自然基底 $\{e_1, e_2, \cdots, e_n\}$ 便是它们的一个标准正交基底。

例2. 在具有标准内积的空间 $(\mathbf{R}^3, (\cdot|\cdot))$ 中，用 G-S 过程由基底

$$x_1=(3, 0, 4)^T, \quad x_2=(-1, 0, 7)^T,$$
$$x_3=(2, 9, 11)^T$$

求标准正交基底的计算是：

$$y_1=(3, 0, 4)^T, \quad \|y_1\|=5,$$

$$z_1=\left(\frac{3}{5}, 0, \frac{4}{5}\right)^T;$$

$$y_2=(-1, 0, 7)^T$$
$$-(-1, 0, 7)\left(\frac{3}{5}, 0, \frac{4}{5}\right)^T\left(\frac{3}{5}, 0, \frac{4}{5}\right)^T$$
$$=(-4, 0, 3)^T$$

$$\|y_2\|=5, \quad z_2=\left(-\frac{4}{5}, 0, \frac{3}{5}\right)^T;$$

$$y_3=(2,9,11)^T$$
$$-(2,9,11)\left(\frac{3}{5},0,\frac{4}{5}\right)^T\left(\frac{3}{5},0,\frac{4}{5}\right)^T$$
$$-(2,9,11)\left(-\frac{4}{5},0,\frac{3}{5}\right)^T\left(-\frac{4}{5},0,\frac{3}{5}\right)^T$$
$$=(0,9,0)^T,$$

$\|y_3\|=9$, $z_3=(0,1,0)^T$

最后得到：$\left\{\left(\frac{3}{5},0,\frac{4}{5}\right)^T,\left(-\frac{4}{5},0,\frac{3}{5}\right)^T,(0,1,0)^T\right\}$

例3. 在一的例3中的内积空间$(\mathbf{C}^{n\times n},(\cdot|\cdot))$及$(\mathbf{R}^{n\times n},(\cdot|\cdot))$中，基底$\{E_{ij}\}$，$i$，$j=1,2,\cdots,n$便是一个标准正交基底。因为，若$\{e_1,e_2,\cdots,e_n\}$是$\mathbf{R}^n$中的自然基底，则有$E_{ij}=e_ie_j^T$，从而有

$$
\begin{aligned}
(E_{ij}|E_{kl})&=\mathrm{tr}(E_{kl}^HE_{ij})=\mathrm{tr}(e_le_k^Te_ie_j^T)\\
&=\mathrm{tr}(\delta_{ki}e_le_j^T)=\mathrm{tr}(\delta_{ki}E_{lj})\\
&=\begin{cases}1, & k=i\text{且}l=j\\ 0, & \text{其它}\end{cases}
\end{aligned}
$$

例4. 在按2.1例4所定义的内积空间$R[x]_1$中，由基底$\{1,x\}$计算出一个标准正交基底的步骤是

$$y_1=1,\ \|y\|^2=\int_0^1 1dx=1,\ z=y_1/\|y_1\|=1$$

$$y_2=x-(x|z_1)z_1=x-\int_0^1 x\cdot 1dx=x-\frac{1}{2}$$

$$\|y_2\|^2=\int_0^1\left(x-\frac{1}{2}\right)^2dx=\frac{1}{12},\ \|y_2\|=\frac{1}{2\sqrt{3}}$$

$$z_2=y_2/\|y_2\|=\sqrt{3}(2x-1)。$$

$\{1,\sqrt{3}(2x-1)\}$便是$R[x]_1$的一个标准正交基底。

四、正交补与投影定理

设W是内积空间V的一个非空子集，x是V中的一个向量。如果x与W中的每个向量都正交，则记之为$x\perp W$。对于V中的两个非空子集W_1，W_2，如果任取$x\in W_1$，$y\in W_2$都有$(x|y)=0$，即$x\perp y$，则称W_1与W_2是互相正交的，记之为$W_1\perp W_2$。关于子集W，可作集合

$$W^{\perp}=\{x\in V \mid x\perp W\},$$

很容易证明它是V的一个线性子空间。

定义1.2.10 设W_1，W_2是内积空间V的两个子空间，且满足

1. $V=W_1+W_2$，
2. $W_1\perp W_2$

则称V有一个正交直和分解，并把W_2叫做W_1的正交补，记之为$W_2=W_1^{\perp}$。从而有$V=W_1+W_1^{\perp}$。

从上述定义中的条件2，很容易推出$W_1\cap W_2=\{0\}$，所以W_1+W_2是直和。又$W_1\perp W_2$，故叫做正交直和。象§1中的定理1.1.22一样，在内积空间中有如下定理。

定理1.2.11 设W为n维内积空间V的一个子空间，则一定存在W的正交补空间$W^{\perp}$，使得$V=W+W^{\perp}$。

证明 设$\dim W=k$，且$\{x_1, x_2, \cdots, x_k\}$为$W$的一个标准正交基底，适当地扩充可得到$V$的标准正交基底$\{x_1,x_2,\cdots,x_k, x_{k+1}, \cdots, x_n\}$。则显然有

$$W^{\perp}=\text{span}(x_{k+1}, x_{k+2}, \cdots, x_n)$$

从而有$V=W+W^{\perp}$。

这个定理表明：有限维内积空间中的每个子空间W，皆有正交补空间$W^{\perp}$存在，且

$$V=W+W^{\perp}$$

而每个$x\in V$有唯一的表示

$$x=u+v,\ u\in W,\ v\in W^{\perp},$$

我们把u叫做x的沿着空间$W^{\perp}$向W上的正交投影，而v叫做x的沿W向$W^{\perp}$上的正交投影。

定义1.2.12 设M为内积空间V中的非空子集，x为V中给定的向量，满足下述等式

$$\|x-m_0\|=\inf_{m\in M}\|x-m\|$$

的$m_0\in M$，叫做x在M上的最佳逼近。

定理1.2.13（最佳逼近的特征） 设M为内积空间V的子空间，x为V中的向量。则$m_0\in M$为x在M上的最佳逼近的充要条件是$x-m_0\perp M$。

证明 必要性。设$m_0\in M$为x在M上的最佳逼近，且$x-m_0$不正交于M，则在M中至少有$m_1\neq 0$且$\|m_1\|=1$，使得$(x-m_0|m_1)=\delta\neq 0$。作$m_2=m_0+\delta m_1\in M$，则有

$$\begin{aligned}\|x-m_2\|^2&=(x-m_0-\delta m_1|x-m_0-\delta m_1)\\&=\|x-m_0\|^2-|\delta|^2\end{aligned}$$

由于$|\delta|^2\neq 0$，故$\|x-m_2\|<\|x-m_0\|$。这与$m_0\in M$是x在M上的最佳逼近相矛盾。故有$x-m_0\perp M$。

充分性。若$x-m_0\perp M$，则任取$m\in M$都有

$$\begin{aligned}\|x-m\|^2&=\|x-m_0+m_0-m\|^2\\&=(x-m_0+m_0-m|x-m_0+m_0-m)\\&=\|x-m_0\|^2+\|m_0-m\|^2\\&\geqslant\|x-m_0\|^2\end{aligned}$$

这表明$m_0\in M$是x在M上的最佳逼近。

这个定理给出了最佳逼近的特征：如果x在子空间M上有最佳逼近m_0的话，则必定满足$x-m_0\perp M$；反之，若M中有m_0满足$x-m_0\perp M$，则它一定是x在M上的最佳逼近。这个特征给出了求最佳逼近的方法。然而，这个定理并没有解决最佳逼近的存在性。如果存在的话，则很容易证明是唯一的。

定理1.2.14（最佳逼近的存在性及计算方案） 设M是内积空间V中的一个k维子空间，则V中的任一向量x都在M上有唯一的最佳逼近。

证明 我们通过具体构造来证明最佳逼近的存在与唯一。设$\{x_1, x_2, \cdots, x_k\}$是$M$的一个基底，$m_0\in M$为$x$在$M$上的最佳逼近。则首先有

$$m_0=\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_kx_k,$$

据定理1.2.13，$x-m_0\perp M$，则更有

$$x-m_0\perp x_i,\ i=1,2,\cdots,k$$

从而得到方程组

$$(x-m_0|x_j)=(x|x_j)-\sum_{i=1}^{k}\lambda_i(x_i|x_j)=0$$

$$j=1,2,\cdots,k,$$

再作整理，便得到方程组

$$\begin{pmatrix}(x_1|x_1) & (x_2|x_1) & \cdots & (x_k|x_1)\\ (x_1|x_2) & (x_2|x_2) & \cdots & (x_k|x_2)\\ & & \vdots & \\ (x_1|x_k) & (x_2|x_k) & \cdots & (x_k|x_k)\end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_k\end{pmatrix}=\begin{pmatrix}(x|x_1)\\ (x|x_2)\\ \vdots\\ (x|x_k)\end{pmatrix}$$

如果记$G=G(x_1,x_2,\cdots,x_k)$，$z=(\lambda_1,\lambda_2,\cdots,\lambda_k)^T$

$$b=((x|x_1),(x|x_2),\cdots,(x|x_k))^T$$

则上述方程组的矩阵记法为$G^Tz=b$。由于基底$\{x_1,x_2,\cdots,x_k\}$的格拉姆矩阵是正定的 Hermite 矩阵，所以方程组$G^Tz=b$有唯一的解z，从而存在唯一的最佳逼近$m_0=\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_kx_k$。

从定理的证明过程可以看出：当取的M的基底$\{x_1,x_2,\cdots,x_k\}$是一个标准正交基底时，由于其格拉姆矩阵为单位阵，故$z=b$。从而最佳逼近有一个很简单的表示

$$m_0=\sum_{i=1}^{k}(x|x_i)x_i$$

这表明，m_0是x在诸x_i方向上的正交投影之和，有与解析几何一样的几何解释。欲求m_0，只要先计算$(x|x_i)$，$i=1,2,\cdots,k$，然后代入上式就行了，不必解方程组。

读者不要小看了这个定理，它是很多优化问题，特别是最小二乘问题中常用的一条根本原理，借助它能解决大量的工程实际问题。

也請读者考虑，如果子空间M是无穷维的，上述证明方法会

遇到什么问题？虽然M是有限维的，但基底不易求出，而却知道它是由一组给定的向量$\{y_1, y_2, \cdots, y_s\}$线性张成的，上述的计算程序会在什么地方出现麻烦？

以后我们要经常在欧氏空间或酉空间中考虑问题，所遇到的与之有关的概念和结果，到时再加以讨论和给出。

第三节 綫性变换

线性变换，有时又叫线性映射或线性算子，它是线性空间之间的一种最简单同时也是最重要的变换。不仅线性代数与泛函分析把它作为研究的主要对象，就是其它数学分支和工程技术也都常把它当作研究问题的主要工具。比如数学中的一些基本运算与变换：求极限，求微分，求积分，坐标变换，拉普拉斯变换，付立叶变换等，作为对函数的变换来看，都满足线性性质，有的就是线性变换的例子。所以各种类型的方程，有些就能归结为对算子方程的研究。这一节，我们就有限维空间之间的线性变换作一般性的介绍，主要是借助于矩阵来表示并刻划它们的各种性质。

一、线性变换的定义及例子

定义1.3.1（映射） 设X，Y是两个集合。如果对每个$x\in X$，都有唯一的$y\in Y$与它对应。这种对应的规律，记之为T，称作为从X到Y的一个映射（或变换）。通常把上述定义简记为

$$T: X\longrightarrow Y$$
$$x\longmapsto y$$

在这种情况下，又把$x\longmapsto y$记为$y=Tx$；并把y叫做x在映射T之下的象，而x叫做y的原象。为了叙述的方便，把Tx叫做关于x取象。

X叫做T的定义域，记之为$\mathscr{D}(T)$；集合

$$\mathscr{R}(T)=\{y\in Y \mid y=Tx,\ x\in\mathscr{D}(T)\}$$

叫做T的值域，有时就记为$T(X)$。如果$T(X)=Y$，则称T是一个满射，或说成是映上的；如果对任意不同的x_1，$x_2\in X$有$Tx_1\neq Tx_2$，则称T是一对一的映射；如果T既是一对一的，又是满射，则称T是一一到上的映射，或X，Y之间的一个一一对应。如果T是一一对应，它有逆映射存在：

$$\begin{aligned} T^{-1}:\ & Y \longrightarrow X \\ & y \longmapsto x \end{aligned}$$

其中$y=Tx$。

定义1.3.2（线性映射） 设X，Y是同一个域F上的线性空间，T：$X\longrightarrow Y$是一个从X到Y的映射。如果对任意的x_1，$x_2\in X$及任意的λ，$\mu\in F$恒有

$$T(\lambda x_1+\mu x_2)=\lambda Tx_1+\mu Tx_2,$$

其中等号左端的线性运算是X中所定义的运算，而等号右端的线性运算是Y中所定义的运算，则称T是从X到Y的一个线性映射或线性变换。

集合

$$\mathscr{N}(T)=\{x\in X \mid Tx=0\}$$

$$\mathscr{R}(T)=\{y\in Y \mid y=Tx,\ x\in X\}$$

分别叫做线性变换T的核和值域。根据线性变换T的定义，读者不难验证，$\mathscr{N}(T)$是X的一个线性子空间，又叫做T的零空间，其维数叫做T的零度；$\mathscr{R}(T)$是Y的一个线性子空间，其维数叫做T的秩。

例1. 如果V是一个线性空间，映射

$$\begin{aligned} T:\ & V \longrightarrow V \\ & x \longmapsto x \end{aligned}$$

显然是一个线性变换，这时记$T=I$，叫做恒等变换。映射

$$\begin{aligned} T:\ & V \longrightarrow V \\ & x \longrightarrow 0 \end{aligned}$$

也是一个线性变换，这时记$T=0$，叫做零变换。

例2. 设$R[x]$是所有实系数多项式构成的实线性空间，$D=\frac{d}{dx}$，则

$$D: R[x] \longrightarrow R[x]$$
$$f(x) \longmapsto Df(x)=f'(x)$$

是一个线性变换，其中$f(x)=a_0x^k+a_1x^{k-1}+\cdots+a_{k-1}x+a_k$，$Df(x)=f'(x)=ka_0x^{k-1}+(k-1)a_1x^{k-2}+\cdots+a_{k-1}$。

例3. 给定$A\in \mathbf{R}^{m\times n}$，映射

$$T: \mathbf{R}^n \longrightarrow \mathbf{R}^m$$
$$x \longmapsto Ax$$

是一个线性变换。

例4. 给定$P\in \mathbf{C}^{m\times m}$，$Q\in \mathbf{C}^{n\times n}$，映射

$$T: \mathbf{C}^{m\times n} \longrightarrow \mathbf{C}^{m\times n}$$
$$X \longmapsto PXQ$$

是一个线性变换。

例5. 设$C[a, b]$表示定义在$[a, b]$上的所有实值连续函数构成的实线性空间，$C^1[a, b]$表示定义在$[a, b]$上所有实值连续可微函数构成的实线性空间。则映射

$$T: C[a, b] \longrightarrow C^1[a, b]$$
$$f(x) \longmapsto \int_a^x f(t)dt$$

其中$x\in[a, b]$，是一个线性变换。

以上的例子，请读者自行验证。

二、线性变换的几个简单性质

1. $T(0)=0$。因为$T(0)=T(0+0)=T(0)+T(0)$。也就是说，线性变换把零向量变成零向量。

2． $T\left(\sum_{i=1}^{n}\lambda_i x_i\right)=\sum_{i=1}^{n}\lambda_i T x_i$。也就是说，任一组向量的线性组合取象，等于分别取象再线性组合。

3． 一组线性相关的向量$\{x_1, x_2, \cdots, x_n\}$，其在线性变换$T$之下的象$\{Tx_1, Tx_2, \cdots, Tx_n\}$仍是线性相关的。这一性质可直接从前两性质推出。

值得注意的是，线性无关的向量组在线性变换之下的象未必线性无关，这要取决于具体的线性变换。

4． 设$T: X \longrightarrow Y$是一个线性变换，且X为有限维空间，$\mathscr{N}(T)$，$\mathscr{R}(T)$分别是T的核及值域。则有

$$\dim X=\dim\mathscr{N}(T)+\dim\mathscr{R}(T)$$

证明 设$\dim X=n$，$\dim\mathscr{N}(T)=k$，$\{x_1, x_2, \cdots, x_k\}$为$\mathscr{N}(T)$的一个基底，$\{x_1, x_2, \cdots, x_k, x_{k+1}, \cdots, x_n\}$为$X$的一个基底。则可证明$\{Tx_{k+1}, Tx_{k+2}, \cdots, Tx_n\}$是$\mathscr{R}(T)$的一个基底。因为任取$x\in X$有

$$x=\lambda_1 x_1+\lambda_2 x_2+\cdots+\lambda_n x_n$$

$$Tx=\lambda_1 Tx_1+\cdots+x_k Tx_k+\lambda_{k+1}Tx_{k+1}+\cdots+\lambda_n Tx_n$$

由于当$x_i\in\mathscr{N}(T)(1\leqslant i\leqslant k)$时有$Tx_i=0$，故又有

$$Tx=\lambda_{k+1}Tx_{k+1}+\cdots+\lambda_n Tx_n$$

这说明$\mathscr{R}(T)$中的每个向量Tx都可由$\{Tx_{k+1}, \cdots, Tx_n\}$线性表出。为了证明$\{Tx_{k+1}, \cdots, Tx_n\}$是线性无关的，设有标量$\lambda_i\in F$，$k+1\leqslant i\leqslant n$，使得

$$\sum_{i=k+1}^{n}\lambda_i Tx_i=0$$

据性质2有

$$T\left(\sum_{i=k+1}^{n}\lambda_i x_i\right)=0$$

这意味着$\sum_{i=k+1}^{n}\lambda_i x_i\in\mathscr{N}(T)$。故存在标量$\delta_i\in F$，$1\leqslant i\leqslant k$，使

得

$$\sum_{i=k+1}^{n} \lambda_i x_i = \sum_{j=1}^{k} \delta_j x_j$$

由于 $\{x_1, x_2, \cdots, x_n\}$ 是 X 的基底，故有 $\lambda_i = 0$，$k+1 \leqslant i \leqslant n$，$\delta_j = 0$，$1 \leqslant j \leqslant k$。这便证明了性质 4 。

三、线性变换的矩阵表示及线性空间的同构

对于有限维空间上的线性变换，可以用矩阵表示它们。因此，对线性变换的研究便转换到对矩阵的研究。

设 X，Y 是域 F 上的两个线性空间，$\dim X = n$，$\dim Y = m$，$\mathscr{B}_1 = \{x_1, x_2, \cdots, x_n\}$，$\mathscr{B}_2 = \{y_1, y_2, \cdots, y_m\}$ 分别是 X 与 Y 的基底。又设 $T: X \longrightarrow Y$ 是一个线性变换。则每个 Tx_j $(j=1, 2, \cdots, n)$ 都可唯一地用 $\mathscr{B}_2$ 线性表出，不妨设

$$T\alpha_j = \alpha_{1j} y_1 + \alpha_{2j} y_2 + \cdots + \alpha_{mj} y_m, \quad 1 \leqslant j \leqslant n$$

也就是说 $(\alpha_{1j}, \alpha_{2j}, \cdots, \alpha_{mj})^T \in F^m$ 是 $T\alpha_j$ 关于基底 $\mathscr{B}_2$ 的坐标。如果要用矩阵的记法，便有

$$(Tx_1 Tx_2 \cdots Tx_n) = (y_1 y_2 \cdots y_m)\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{pmatrix}$$

如果进一步简化记法，令 $T\mathscr{B}_1 = (Tx_1 Tx_2 \cdots Tx_n)$，$A = (\alpha_{ij}) \in F^{m \times n}$，则上述等式为

$$T\mathscr{B}_1 = \mathscr{B}_2 A$$

这时，我们把矩阵 A 叫做线性变换 T 关于基底 $\mathscr{B}_1$，$\mathscr{B}_2$ 的一个矩阵表示，并记之为

$$m(T)_{\mathscr{B}_1, \mathscr{B}_2} = A$$

特别是，当 $X = Y$ 时，取 $\mathscr{B}_1 = \mathscr{B}_2$，则有

$$T\mathscr{B}_1 = \mathscr{B}_1 A, \quad A = (\alpha_{ij}) \in F^{n \times n},$$

这时，我们把矩阵A叫做线性变换T关于基底$\mathscr{B}_1$的一个矩阵表示，并记之为

$$m(T)_{\mathscr{B}_1}=A$$

综上所述，要想求出线性变换T：$X\longrightarrow Y$的一个矩阵表示A，只要按下面步骤进行：

第一，分别指定X，Y的基底：$\mathscr{B}_1=\{x_1, x_2,\cdots,x_n\}$，$\mathscr{B}_2=\{y_1, y_2, \cdots, y_m\}$；

第二，根据T的定义，计算出Tx_1，Tx_2，…，Tx_n；

第三，求出每个Tx_j关于基底$\mathscr{B}_2$的唯一的坐标$a_j=(a_{1j}, a_{2j}, \cdots, a_{mj})^T\in F^m$；

第四，写出T关于$\mathscr{B}_1$，$\mathscr{B}_2$的矩阵表示

$$m(T)_{\mathscr{B}_1,\mathscr{B}_2}=A=(a_1a_2\cdots a_n)\in F^{m\times n}$$

如果是求线性变换T：$X\longrightarrow X$的一个矩阵表示，则

第一，指定X的一个基底$\mathscr{B}_1=\{x_1, x_2, \cdots, x_n\}$；

第二，根据T的定义，计算出Tx_1，Tx_2，…，Tx_n；

第三，求出每个Tx_j关于基底$\mathscr{B}_1$的唯一的坐标$a_j=(a_{1j}, a_{2j}, \cdots, a_{nj})^T\in F^n$

第四，写出T关于基底$\mathscr{B}_1$的矩阵表示

$$m(T)_{\mathscr{B}_1}=A=(a_1a_2\cdots a_n)\in F^{n\times n}$$

如此看来，求线性变换T的矩阵表示是不难的，只是在第三步的计算中，往往会遇到一些麻烦，或计算工作量过大。下面举几个简单的例子，以说明求矩阵表示的过程。

例1. 求线性变换

$$D: R[x]_n\longrightarrow R[x]_n$$
$$f(x)\longmapsto f'(x)$$

关于基底$\mathscr{B}=\{1, x, x^2, \cdots, x^n\}$的矩阵表示。

解 $D1=0$，$Dx=1$，…，$Dx^n=nx^{n-1}$；$D1$关于$\mathscr{B}$的坐标

为 $(0, 0, \cdots, 0)^T$，Dx关于$\mathscr{B}$的坐标为 $(1, 0, \cdots, 0)^T$，…，Dx^n关于$\mathscr{B}$的坐标为 $(0, 0, \cdots, n, 0)^T$，故有

$$m(D)_{\mathscr{B}} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & n \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbf{R}^{(n+1)\times(n+1)}$$

例2. 求线性变换

$$T: R[x]_n \longrightarrow R[x]_{n+1}$$

$$f(x) \longmapsto \int_0^x f(t)dt$$

关于基底$\mathscr{B}_1 = \{1, x, \cdots, x^n\}$，$\mathscr{B}_2 = \{1, x, \cdots, x^{n+1}\}$的矩阵表示。

解 $T1 = \int_0^x 1dt = x$，$Tx = \int_0^x tdt = \frac{1}{2}x^2$，…，$Tx^n = \int_0^x t^n dt = \frac{1}{n+1}x^{n+1}$；$T1$ 关于 $\mathscr{B}_2$ 的坐标为 $(0, 1, 0, \cdots, 0)^T$，Tx关于$\mathscr{B}_2$的坐标为$\left(0, 0, \frac{1}{2}, 0, \cdots, 0\right)^T$，…，$Tx^n$关于$\mathscr{B}_2$的坐标为$\left(0, 0, \cdots, 0, \frac{1}{n+1}\right)^T$；故有

$$m(T)_{\mathscr{B}_1, \mathscr{B}_2} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \frac{1}{n+1} \end{pmatrix} \in \mathbf{R}^{(n+2)\times(n+1)}$$

例3. 若线性变换T：$\mathbf{R}^2 \longrightarrow \mathbf{R}^3$被定义为：

$$Tx=y$$

其中$x=(\xi_1, \xi_2, \xi_3)^T$，$y=(3\xi_1+\xi_3, 2\xi_1+\xi_2, \xi_1+\xi_2+\xi_3)^T$；试求$T$关于自然基底$\mathscr{B}=\{e_1, e_2, e_3\}$的矩阵表示。

解 $Te_1=(3,2,1)^T$，$Te_2=(0,1,1)^T$，$Te_3=(1,0,1,)^T$据$T\mathscr{B}=\mathscr{B}A$，而$\mathscr{B}=(e_1e_2e_3)=I$，故有

$$m(T)_{\mathscr{B}}=A=T\mathscr{B}=(Te_1Te_2Te_3)=\begin{bmatrix}3 & 0 & 1\\ 2 & 1 & 0\\ 1 & 1 & 1\end{bmatrix}$$

在指定了空间X和Y的基底之后，便可以求线性变换T：$X\longrightarrow Y$关于这对基底的矩阵表示。而空间X，Y的基底一般都不是唯一的（除非它们是仅由零向量构成的空间），自然会提出以下几个问题：

1. 求出了T关于基底$\mathscr{B}_1$，$\mathscr{B}_2$的矩阵表示A有什么用？

2. 如果改指定X，Y的基底为$\mathscr{B}_1'$，$\mathscr{B}_2'$，又可得到T的矩阵表示$B=m(T)_{\mathscr{B}_1', \mathscr{B}_2'}$，试问$A$与$B$之间有什么关系？

3. 如何选定X，Y的基底，方使得T的矩阵表示最为简单？

4. 指定了$\mathscr{B}_1$，$\mathscr{B}_2$之后，线性变换T：$X\longrightarrow Y$有唯一的矩阵表示$m(T)_{\mathscr{B}_1, \mathscr{B}_2}=A$；反过来，如果给定一个矩阵$A\in F^{m\times n}$，是否存在唯一的线性变换$T$：$X\longrightarrow Y$，使得$m(T)_{\mathscr{B}_1, \mathscr{B}_2}=A$？

对于以上问题，我们在后面逐一回答。

对于X中的任一向量$x=\xi_1x_1+\xi_2x_2+\cdots+\xi_nx_n$，则有

$$\begin{aligned}Tx&=T\left(\sum_{j=1}^{n}\xi_jx_j\right)=\sum_{j=1}^{n}\xi_jTx_j\\ &=\sum_{j=1}^{n}\xi_j\sum_{i=1}^{m}\alpha_{ij}y_i=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\alpha_{ij}\xi_j\right)y_i\end{aligned}$$

如果记$a=(\xi_1, \xi_2, \cdots, \xi_n)^T \in F^n$，则上面推导为

$$Tx=T\mathscr{B}_1 a=\mathscr{B}_2 Aa$$

这表明x在T之下的象Tx，关于基底$\mathscr{B}_2$的坐标为Aa；一旦求出了$A=m(T)_{\mathscr{B}_1;\mathscr{B}_2}$，又给出了$x\in X$关于基底$\mathscr{B}_1$的坐标$a$，则其象$Tx$关于基底$\mathscr{B}_2$的坐标便是$Aa$。因此$T$在整个$X$上的作用，由它在基底$\mathscr{B}_1$上的作用所完全确定，而矩阵表示$A$就是刻画$T$作用在$\mathscr{B}_1$上的情况。这样便回答了第一个问题。

如果我们改变指定的X，Y的基底分别为$\mathscr{B}_1'$，$\mathscr{B}_2'$，且T：$X\longrightarrow Y$关于$\mathscr{B}_1'$，$\mathscr{B}_2'$的矩阵表示为B，即$m(T)_{\mathscr{B}_1',\mathscr{B}_2'}=B$；并记$\mathscr{B}_1'=\mathscr{B}_1 P$，$\mathscr{B}_2'=\mathscr{B}_2 Q$，其中$P\in F^{n\times n}$，$Q\in F^{m\times m}$分别是基底$\mathscr{B}_1$到$\mathscr{B}_1'$，基底$\mathscr{B}_2$到$\mathscr{B}_2'$的变换阵，当然都是可逆的。则据关系式

$$T\mathscr{B}_1=\mathscr{B}_2 A,\ T\mathscr{B}_1'=\mathscr{B}_2' B,\ \mathscr{B}_1'=\mathscr{B}_1 P,$$

$$\mathscr{B}_2'=\mathscr{B}_2 Q$$

很容易推出$\mathscr{B}_2 AP=\mathscr{B}_2 QB$，即

$$\mathscr{B}_2(AP-QB)=0$$

由于$\mathscr{B}_2$是一个基底，故可推出（为什么？）

$$AP-QB=0$$

即

$$B=Q^{-1}AP$$

这表明A与B是相抵的，即$A\cong B$。

如果在上述推导过程中，令$Y=X$，$\mathscr{B}_2=\mathscr{B}_1$，$\mathscr{B}_2'=\mathscr{B}_1'$，$\mathscr{B}_1'=P\mathscr{B}_1$，则据关系式

$$T\mathscr{B}_1=\mathscr{B}_1 A,\ T\mathscr{B}_1'=\mathscr{B}_1' B,\ \mathscr{B}_1'=\mathscr{B}_1 P$$

可推出

$$\mathscr{B}_1 AP=\mathscr{B}_2 PB,\ \text{即}$$

$$\mathscr{B}_2(AP-PB)=0$$

据$\mathscr{B}_2$是一个基底，可知

$$AP-PB=0$$

即

$$B=P^{-1}AP$$

这表明A与B是相似的，即$A\sim B$。

综上所述，线性变换T：$X\longrightarrow Y$关于不同基底偶($\mathscr{B}_1$，$\mathscr{B}_2$)的矩阵表示是相抵系关；而线性变换T：$X\longrightarrow X$关于不同基底的矩阵表示是相似关系。这便回答了第二个问题。

关于第三个问题，是我们本章余下部分所要研究的主要课题，放在后面讨论。先看比较容易解决的第四个问题，并且用下面定理给出回答。

定理1.3.3 设X，Y是域F上的线性空间，且维数分别为n与m；又设$\mathscr{B}_1=\{x_1, x_2, \cdots, x_n\}$，$\mathscr{B}_2=\{y_1, y_2, \cdots, y_m\}$分别为$X$，$Y$的基底。给定$A=(a_{ij})\in F^{m\times n}$，则存在唯一的线性变换$T$：$X\longrightarrow Y$，它关于基底偶$\mathscr{B}_1$，$\mathscr{B}_2$的矩阵表示为$A$，即$m(T)_{\mathscr{B}_1,\mathscr{B}_2}=A$。

证明 任取$x\in X$，设它关于$\mathscr{B}_1$的坐标为

$$a=(\xi_1, \xi_2, \cdots, \xi_n)^T\in F^n$$

即$x=\mathscr{B}_1a=\sum_{i=1}^{n}\xi_i x_i$。又取

$$y=\mathscr{B}_2Aa=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}\xi_j\right)y_i\in Y$$

作映射

$$T: X\longrightarrow Y$$
$$x\longmapsto y$$

亦即$Tx=\mathscr{B}_2Aa$。则可以证明T就是定理中所要求的线性变换。

首先证明T是线性的。为此，任取u，$v\in X$，它们关于$\mathscr{B}_1$的坐标分别为

$$p=(\eta_1, \eta_2, \cdots, \eta_n)^T, \quad q=(\gamma_1, \gamma_2, \cdots, \gamma_n)^T$$

即有$u=\mathscr{B}_1p$，$v=\mathscr{B}_1q$；又任取λ，$\mu\in F$，则显然有

$$\lambda u+\mu v=\lambda \mathscr{B}_1 p+\mu \mathscr{B}_1 q=\mathscr{B}_1(\lambda p+\mu q)$$

而据T的定义有$Tu=\mathscr{B}_2 Ap$，$Tv=\mathscr{B}_2 Aq$，且

$$\begin{aligned}T(\lambda u+\mu v)&=\mathscr{B}_2 A(\lambda p+\mu q)\\&=\lambda \mathscr{B}_2 Ap+\mu \mathscr{B}_2 Aq\\&=\lambda Tu+\mu Tv\end{aligned}$$

这表明T是线性的。

再证$T\mathscr{B}_1=\mathscr{B}_2 A$。显然每个$x_i\in\mathscr{B}_1$关于$\mathscr{B}_1$的坐标为$e_i\in\mathbf{R}^n$，即$x_i=\mathscr{B}_1 e_i$。故据$T$的定义有$Tx_i=\mathscr{B}_2 Ae_i$，从而有

$$\begin{aligned}T\mathscr{B}_1&=(Tx_1 Tx_2\cdots Tx_n)\\&=(\mathscr{B}_2 Ae_1 \mathscr{B}_2 Ae_2\cdots \mathscr{B}_2 Ae_n)\\&=\mathscr{B}_2 A(e_1 e_2\cdots e_n)=\mathscr{B}_2 AI=\mathscr{B}_2 A\end{aligned}$$

最后证明T的唯一性。若还有线性变换T_1：$X\longrightarrow Y$也满足$T_1\mathscr{B}_1=\mathscr{B}_2 A$，则把该式与

$$T\mathscr{B}_1=\mathscr{B}_2 A$$

相减，便得到$T_1\mathscr{B}_1-T\mathscr{B}_1=0$，即$T_1\mathscr{B}_1=T\mathscr{B}_1$，亦即对每个$x_i\in\mathscr{B}_1$都有$T_1 x_i=Tx_i$。从而对任意的$x\in X$，且设$x=\sum_{i=1}^{n}\xi_i x_i$，都有

$$T_1 x=\sum_{i=1}^{n}\xi_i T_1 x_i=\sum_{i=1}^{n}\xi_i Tx_i=Tx$$

这表明T_1与T作用在X上完全是一致的，故有$T=T_1$。

上面定理告诉我们：在指定了X，Y的一对基底之后，每个线性变换T：$X\longrightarrow Y$关于这对基底有唯一的矩阵表示A；反过来，每个矩阵$A\in F^{m\times n}$也能唯一地定义一个线性变换T：$X\longrightarrow Y$，且它关于指定的基底的矩阵表示恰好是A。在这种意义下，矩阵空间$F^{m\times n}$可以与集合

$$L(X, Y)=\{T\mid T\text{是}X\text{到}Y\text{的线性变换}\}$$

建立起一一对应的关系。如果对集合$L(X, Y)$定义如下运算：

$\forall T, S \in L(X, Y), \lambda, \mu \in F, x \in X$,

$$(T+S)x \triangleq Tx + Sx$$

$$(\lambda T)x \triangleq \lambda Tx$$

则不难验证$L(X, Y)$是域F上的一个线性空间。并且可以证明线性空间$F^{m\times n}$与$L(X, Y)$在代数上来看有完全相同的结构，只是各自所包含的元素不同而已。

定义1.3.4 同一个域F上的线性空间X与Y之间，如果存在一个映射σ：$X \longrightarrow Y$，对一切$x_1, x_2 \in X$及$\lambda, \mu \in F$满足：

1. $\sigma(X)=Y$，σ是到上的映射；
2. 当$x_1 \neq x_2$时有$\sigma(x_1) \neq \sigma(x_2)$，$\sigma$是一对一的；
3. $\sigma(\lambda x_1 + \mu x_2) = \lambda\sigma(x_1) + \mu\sigma(x_2)$，$\sigma$是线性的；

则称X与Y是（代数上）同构的，而σ叫做X与Y之间的一个同构映射。换句话说，如果在X，Y之间存在一个一一到上的线性变换σ，则称X与Y是同构的线性空间。

定理1.3.5 空间$F^{m\times n}$与$L(X, Y)$是同构的。

证明 指定了X与Y的基底$\mathscr{B}_1$与$\mathscr{B}_2$之后，则任一$T \in L(X, Y)$关于这一对基底有唯一的矩阵$A \in F^{m\times n}$与之对应，即

$$m(T)_{\mathscr{B}_1, \mathscr{B}_2} = A \text{或} T\mathscr{B}_1 = \mathscr{B}_2 A$$

现在作映射

$$\sigma: L(X, Y) \longrightarrow F^{m\times n}$$
$$T \longmapsto A$$

则前面的定理1.3.3表明σ是一一到上的映射，所以只须证明σ是线性的。为此，任取$S_1, S_2 \in L(X, Y)$，$\lambda, \mu \in F$，且设$S_1\mathscr{B}_1 = \mathscr{B}_2 C$，$S_2\mathscr{B}_1 = \mathscr{B}_2 D$，$C, D \in F^{m\times n}$，即$\sigma(S_1)=C$，$\sigma(S_2)=D$；则按$\lambda S_1 + \mu S_2 \in L(X, Y)$的定义：对任意的$x \in X$有

$$(\lambda S_1 + \mu S_2)x = \lambda S_1 x + \mu S_2 x$$

故对每个$x_i \in \mathscr{B}_1$，也有

$$(\lambda S_1+\mu S_2)x_i=\lambda S_1 x_i+\mu S_2 x_i$$

从而有

$$\begin{aligned}(\lambda S_1+\mu S_2)\mathscr{B}_1&=\lambda S_1\mathscr{B}_1+\mu S_2\mathscr{B}_1\\&=\lambda\mathscr{B}_2 C+\mu\mathscr{B}_2 D\end{aligned}$$

此即
$$\sigma(\lambda S_1+\mu S_2)=\lambda\sigma(S_1)+\mu\sigma(S_2)$$

这便证明了σ是线性的。

定理1.3.6 设σ是线性空间X，Y之间的一个同构映射，则X中任一组线性无关的向量$\{x_1, x_2, \cdots, x_k\}$在$\sigma$之下的象$\{\sigma x_1, \sigma x_2, \cdots, \sigma x_k\}$也是线性无关的。

证明 设有$\lambda_1, \lambda_2, \cdots, \lambda_k\in F$使得

$$\lambda_1\sigma x_1+\lambda_2\sigma x_2+\cdots+\lambda_k\sigma x_k=0$$

则据σ的线性性，有$\sigma\left(\sum_{i=1}^{k}\lambda_i x_i\right)=0$；又由于$\sigma$是一一到上的线性映射，故只有$\sum_{i=1}^{k}\lambda_i x_i=0$时方有$\sigma\left(\sum_{i=1}^{k}\lambda_i x_i\right)=0$。而由于$\{x_1, x_2, \cdots, x_k\}$是线性无关的，故有$\lambda_1=\lambda_2=\cdots=\lambda_k=0$。

定理1.3.7 每个域F上的n维线性空间V都与F^n同构。从而每个n维的实线性空间都与$\mathbf{R}^n$同构，每个n维的复线性空间都与$\mathbf{C}^n$同构。

证明 设$\mathscr{B}=\{x_1, x_2,\cdots, x_n\}$是$V$的一个指定的基底，则任一$x\in V$都有

$$x=\xi_1 x_1+\xi_2 x_2+\cdots+\xi_n x_n=(x_1 x_2\cdots x_n)a$$

其中$a=(\xi_1, \xi_2, \cdots, \xi_n)^T\in F^n$，且$a$是唯一的。这说明$x\in V$与其（关于基底$\mathscr{B}$的）坐标$a$有一个对应关系，记之为$\sigma$，即$\sigma(x)=a$；很显然，不同的$x$，$y\in V$，关于基底$\mathscr{B}$的坐标是不同的。从而表明映射

$$\begin{aligned}\sigma:\ V&\longrightarrow F^n\\x&\longrightarrow a\end{aligned}$$

是一对一的映射。另外，任取$a\in F^n$，也都有$x=\mathscr{B}a\in V$与之对

应，说明σ是到上的。为了证明σ是线性的，任取x，$y \in V$，λ，$\mu \in F$，且设$x=\mathscr{B}a$，$y=\mathscr{B}b$，即$\sigma(x)=a$，$\sigma(y)=b$。则因

$$\lambda x+\mu y=\lambda \mathscr{B}a+\mu \mathscr{B}b=\mathscr{B}(\lambda a+\mu b)$$

即

$$\sigma(\lambda x+\mu y)=\lambda a+\mu b=\lambda \sigma(x)+\mu\sigma(y)$$

故证明了σ是一个线性映射，从而证明了σ是V与F^n之间的一个同构映射。

推论1.3.8 对于线性空间V中的线性无关组$\{x_1,x_2,\cdots,x_k\}$，其关于V的任一基底$\mathscr{B}$的坐标$\{a_1, a_2, \cdots, a_k\}$在$F^n$中也是线性无关的，反之亦然。

推论1.3.9 设$\mathscr{B}_1$，$\mathscr{B}_2$分别为X，Y的基底，$T \in L(X,Y)$ $m(T)_{\mathscr{B}_1,\mathscr{B}_2}=A=(a_1a_2\cdots a_n)\in F^{m\times n}$；又设

$$\mathscr{R}(T)=\{y\in Y \mid y=Tx,\ x\in X\}$$

$$\mathscr{R}(A)=\{v\in F^m \mid v=Au,\ u\in F^n\}$$

则$\mathscr{R}(T)$与$\mathscr{R}(A)$是同构的，即

$$\dim\mathscr{R}(T)=\dim\mathscr{R}(A)=\mathrm{rank}(A)=\mathrm{rank}(T)$$

证明 不失一般性，我们就这种情况来证明：$T \in L(X,Y)$不是零变换，且指定的基底$\mathscr{B}_1=\{x_1, x_2, \cdots, x_n\}$中的向量的排列满足：

$$Tx_i\neq 0,\ 1\leqslant i\leqslant r;\ Tx_j=0,\ r+1\leqslant j\leqslant n;$$

由§3.2线性变换的性质4的证明知

$$\mathscr{R}(T)=\mathrm{span}(Tx_1, Tx_2, \cdots, Tx_r)$$

且$\{Tx_1, Tx_2, \cdots, Tx_r\}$为$\mathscr{R}(T)$的一个基底。故有

$$\mathrm{rank}(T)=\dim\mathscr{R}(T)=r$$

另一方面，又有

$$\mathscr{R}(A)=\mathrm{span}(a_1, a_2, \cdots, a_n)$$

而据$m(T)_{\mathscr{B}_1,\mathscr{B}_2}=A$的定义知$T\mathscr{B}_1=\mathscr{B}_2A$，即

$$(Tx_1Tx_2\cdots Tx_n)=(y_1, y_2, \cdots, y_m)[a_1a_2\cdots a_n]$$

说明A的n个列向量a_1，a_2，…，a_n分别是Tx_1，Tx_2，…，Tx_n关于基底$\mathscr{B}_2=\{y_1, y_2, \cdots, y_m\}$的坐标。又由于$Y$与$F^m$同构，故由推论1.3.8知，$a_1$，$a_2$，…，$a_r$是线性无关的，且据$Tx_j=0$ $(r+1\leqslant j\leqslant n)$ 知$a_j=0$ $(r+1\leqslant j\leqslant n)$，从而说明

$$\mathscr{R}(A)=\mathrm{span}(a_1, a_2, \cdots, a_r) \text{ 且} \mathrm{rank}(A)=r$$

由于$\mathscr{R}(T)$与$\mathscr{R}(A)$是同一个域F上的r维空间，故是同构的。

四、线性变换的最简矩阵表示及特征值

一个线性变换关于空间的不同基底有不同的矩阵表示，但这些矩阵两两相似。所以一个线性变换能否有最简单的矩阵表示，等价于它的一个矩阵表示在相似的意义下，能有怎样的最简形式。从矩阵的角度来看，零矩阵及单位矩阵最为简单，但它们对应的线性变换分别是零变换及恒等变换，这未免太特殊了。除此而外，纯量矩阵aI对应的是相似变换aI，也比较特殊，所以我们考虑对角矩阵。

如果线性变换T关于空间的某一基底$\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$有矩阵表示

$$m(T)_{\mathscr{B}}=D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$$

则显然有

$$Tx_i=d_ix_i, \qquad i=1, 2, \cdots, n$$

如果D的主对角元$\{d_1, d_2, \cdots, d_n\}$中共有$r$个不为零，不妨设为$d_{i_1}, d_{i_2}, \cdots, d_{i_r}$，其余的$n-r$个为零，设为$d_{j_1}$，$d_{j_2}, \cdots, d_{j_{n-r}}$，则很容易看出

$$\mathscr{R}(T)=\mathrm{span}(x_{i_1}, x_{i_2}, \cdots, x_{i_r})$$

$$\mathscr{N}(T)=\mathrm{span}(x_{j_1}, x_{j_2}, \cdots, x_{j_{n-r}})$$

其中x_{i_k}，x_{j_l}为基底$\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$中的向量。所以$T$的秩与零度是一目了然的。而且线性变换的其它数值特征，如特征值等也是显然的。但是，还必须考虑这样的问题：

第一，每个线性变换T都能有对角矩阵表示吗？或者说，怎样的T才有对角阵表示？

第二，如果T有对角阵表示，那么关于哪一个基底才有对角阵表示？如何求它？

第三，如果T沒有对角阵表示，那么它的最简矩阵表示又具有什么样的形式？

以上问题是我们要深入讨论的线索。为此我们先从T的特征值和特征向量入手。

定义1.3.10 设X是域F上的n维线性空间，而$T: X \longrightarrow X$是一个线性变换。如果有$\lambda \in F$，$x \in X$且$x \neq 0$，使得$Tx = \lambda x$，则称λ为T的一个特征值，而x为T关于λ的特征向量。而所有满足方程$Tx = \lambda x$的向量x的集合，构成X的一个线性子空间，记之为$E(\lambda)$，叫做T关于λ的特征子空间。

如果T关于X的基底$\mathscr{B} = \{x_1, x_2, \cdots, x_n\}$有对角矩阵$D$表示，即$m(T)_{\mathscr{B}} = D = \operatorname{diag}(d_1, \cdots, d_n)$，

$$(Tx_1, Tx_2, \cdots, Tx_n) = (x_1, x_2, \cdots, x_n)D$$
$$= (d_1x_1, d_2x_2, \cdots, d_nx_n)$$

从而可知

$$Tx_i = d_ix_i, \qquad i = 1, 2, \cdots, n$$

这表明每个基向量x_i都是T关于其特征值d_i的特征向量。从而进一步得到这样的结论：

T有对角阵表示的必要条件是T有n个线性无关的特征向量。

这个条件是否充分呢？如果T有n个性线无关的特征向量$x_1, x_2, \cdots, x_n$，其对应的特征值为$\lambda_1, \lambda_2, \cdots, \lambda_n$（它们可以相同），即

$$Tx_i = \lambda_ix_i, \quad i = 1, 2, \cdots, n$$

则$\{x_1, x_2, \cdots, x_n\}$可构成$X$的一个基底$\mathscr{B}$，很显然有

$$
\begin{aligned}
&(Tx_1, Tx_2, \cdots, Tx_n) \\
&=(\lambda_1 x_1, \lambda_2 x_2, \cdots, \lambda_n x_n) \\
&=(x_1, x_2, \cdots, x_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.
\end{aligned}
$$

即 $T\mathscr{B}=\mathscr{B}\Lambda$，其中 $\Lambda=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, T_n)$。这表明 $m(T)_{\mathscr{B}}=\Lambda$ 是一个对角阵。从而说明条件也是充分的。这样便得到上面第一个问题的回答：

定理1.3.11 线性变换 $T: X \longrightarrow X$：有对角阵表示的充分必要条件是它有 n 个线性无关的特征向量。

要想用上面的定理来判断 T 是否有对角阵表示，就必须求出 T 的所有特征向量，然后检验是否有 n 个线性无关的。找出了 n 个线性无关的特征向量，便可作为 X 的基底，T 关于这个基底的矩阵表示就是以相应的 n 个特征值为主对角元的对角矩阵。一般来说，按特征向量的定义求解算子方程 $Tx=\lambda x$ 是难以达到目的的。为此，还是借助于 T 的矩阵表示来考虑问题。

任指定 X 的一个基底 $\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$，求出 T 关于 $\mathscr{B}$ 的矩阵表示 $m(T)_{\mathscr{B}}=A$。设 $x \in X$ 是 T 的一个特征向量，λ 是相应的特征值，即

$$
Tx=\lambda x, \qquad x \neq 0, \ \lambda \in F
$$

则 x 可用 $\mathscr{B}$ 线性表出，即

$$
x=\sum_{i=1}^{n} \xi_i x_i=\mathscr{B}\xi
$$

其中 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)^T$ 是 x 关于 $\mathscr{B}$ 的坐标，$\xi \in F^n$。从而有

$$
Tx=T\mathscr{B}\xi=\mathscr{B}A\xi, \qquad \lambda x=\mathscr{B}(\lambda\xi)
$$

根据 $Tx=\lambda x$ 有 $\mathscr{B}A\xi=\mathscr{B}(\lambda\xi)$。这样便得到 T 的特征向量 x 关

于$\mathscr{B}$的坐标ξ所满足的方程

$$A\xi=\lambda\xi, \qquad \xi\neq 0, \ \lambda\in F$$

这是一个线性方程组，有与

$$Tx=\lambda x, \qquad x\neq 0, \ \lambda\in F$$

相似的形式。如果我们也把ξ定义为矩阵 A 的特征向量，λ 为相应的特征值，则 T 与 A 有相同的特征值，而在特征向量 x 与 ξ 之间的关系也很清楚，ξ 为 x 关于$\mathscr{B}$的坐标。所以，如果能求出 A 的所有特征值和特征向量，则 T 的所有特征值和特征向量也就可知了。再根据空间 X 与 F^n 是同构的，因为有同构映射

$$\sigma: X\longrightarrow F^n$$
$$x\longmapsto\xi$$

所以检验 A 有无 n 个线性无关的特征向量就够了。因此，关于 T 所提出的问题，就转化为关于其矩阵表示 A 所提出的问题。下面仅就矩阵来考虑我们所提出的问题。

定义1.3.12 设 $A\in F^{n\times n}$，如果存在 $\lambda\in F$，$0\neq x\in F^n$ 满足 $Ax=\lambda x$，则称λ为A的一个特征值，x 为 A 的关于 λ 的特征向量。

方程$(\lambda I-A)x=0$ 要有非零解，则其系数矩阵$\lambda I-A$ 必须是降秩的。所以有下述结论：

定理1.3.13 对于矩阵 $A\in F^{n\times n}$，下述命题是等价的：

1. λ是A的一个特征值，
2. 矩阵$\lambda I-A$是奇异的，
3. $\det(\lambda I-A)=0$

其中（3）比较重要，因为它告诉我们如何去求A的特征值。

把行列式 $\det(\lambda I-A)$ 展开可知它是 λ 的一个 n 次多项式，多项式的系数属于F。而λ就是这个多项式的根（限制在F上求根）。设

$$A=(a_{ij})\in F^{n\times n}$$

则

$$f(\lambda) \hat{=} \det(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ & \vdots & & \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

$$= \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n, \qquad \alpha_i \in F$$

我们把$f(\lambda)$叫做A的特征多项式，A的特征值λ就是方程$f(\lambda)=0$在F中的根。所以，欲求A的特征值，需先求特征多项式$f(\lambda)$，然后解方程$f(\lambda)=0$。一般来说，这是相当困难的，在数值计算上很少沿这一途径去做。

例1. 设T是$\mathbf{R}^2$上的一个线性变换，它关于标准基底$\{e_1, e_2\}$的矩阵表示为$A=\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$，则它的特征多项式为

$$f(\lambda) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 + 1$$

由于$f(\lambda)=\lambda^2-1$在$\mathbf{R}$上沒有根，故T在$\mathbf{R}$上沒有特征值，当然也沒有特征向量，从而T沒有对角阵表示。

如果把T定义在$\mathbf{C}^2$上，其关于$\{e_1, e_2\}$的矩阵表示仍为例1中的A，则$f(\lambda)=\lambda^2+1$在$\mathbf{C}$上有两根i，$-i$，这时T（或A）便有两个特征值，以后会知道T有对角阵表示。所以在讨论特征值问题时，必须小心线性空间所依赖的域。

例2. 设T是$\mathbf{R}^3$上的线性变换，其关于标准基底$\{e_1, e_2, e_3\}$的矩阵表示为

$$A = \begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

则计算可知T（或A）的特征多项式为

$$f(\lambda) = \det(\lambda I - A) = \lambda^3 - 5\lambda^2 + 8\lambda - 4$$
$$= (\lambda - 1)(\lambda - 2)^2$$

故有特征值$\lambda_1=1$，$\lambda_2=\lambda_3=2$。将λ_1，λ_2分别代入$(\lambda I-A)x=0$，则有

$$(\lambda_1 I-A)x=(I-A)x=0$$
$$(\lambda_2 I-A)x=(2I-A)x=0$$

由于$\mathrm{rank}(I-A)=2$，故求得$(I-A)x=0$的基础解系为$x_1=(1,0,2)^T$。T关于λ_1只有一个线性无关的特征向量，即$E(\lambda_1)=\mathrm{span}(x_1)$。由于$\mathrm{rank}(2I-A)=2$，故求得$(2I-A)x=0$的基础解系为$x_2=(1,1,2)^T$，$T$关于$\lambda_2$也只有一个线性无关的特征向量，即$E(\lambda_2)=\mathrm{span}(x_2)$。$T$是三维空间上的线性算子，而只有二个线性无关的特征向量，故没有对角阵表示。也就是说，它的矩阵表示A不能与对角阵相似。这种情况，一般把T（或A）叫做亏损的。很多教科书上又把有对角阵表示的线性变换T，或与对角阵相似的方阵A，叫做是可对角化的。

上面的讨论，给出了求线性变换T的特征向量的一种程序：

1. 指定空间X的一个基底$\mathscr{B}$，求出T关于这个基底的矩阵表示$m(T)_{\mathscr{B}}=A$。

2. 求A（或T）的特征多项式

$$f(\lambda)=\det(\lambda I-A)$$

并在限定的域F上求$f(\lambda)=0$的根：

$$\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k,\qquad 0\leqslant k\leqslant n$$

3. 分别把特征值λ_i代入方程，便得

$$(\lambda_i I-A)x=0\quad i=1,2,\cdots,l$$

其中l为不同特征值的个数。求解这些方程的解，它们是T的特征向量关于基底$\mathscr{B}$的坐标。从而可求得T的特征向量。

4. 通过考查A的所有特征向量的极大线性无关组包含向量的个数，便可判定A能否与对角阵相似，即T能否对角化。

当然，上面的判别法则不是唯一的，也不是最好的，在某些情形之下，还是不可行的。但它是最基本的，很多具有技巧性的

判别法是在此基础上演变而来的。下面我们进一步分析可对角化的线性变换T还有什么特征。

假定T是可对角化的，λ_1，λ_2，…，λ_k是它的所有不同特征值，则存在一个基底$\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$，使得

$$m(T)_{\mathscr{B}}=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

其中λ_i有重复的时，不妨设λ_i的重数为d_i，适当地把$\mathscr{B}$中的基向量调动次序，变为$\mathscr{B}'$，则有

$$m(T)_{\mathscr{B}'}=\begin{pmatrix} \lambda_1 I_1 & & & \\ & \lambda_2 I_2 & & \\ & & \ddots & \\ & & & \lambda_k I_k \end{pmatrix}$$

其中$I_j \in F^{d_j \times d_j}$。从这个矩阵可以看出两个重要的事实：

第一，T的特征多项式$f(\lambda)$是线性因子的幂积，即

$$f(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_k)^{d_k}$$

这表明可对角化的T，其特征多项式一定能分解为一次因式的幂积。所以，如果F是代数上封闭的，比如$F=\mathbf{C}$，则它上面的任一多项式皆可有如此的分解，这时T有可能可对角化。反之，如果F不是代数上封闭的，如$F=\mathbf{R}$，则它上面的多项式不一定能有上述的分解。因此，T的特征多项式若不能作如此分解，则一定是不可对角化的。要特别强调的是，即使T的特征多项式有如此的分解，也未必可对角化。

第二，λ_i作为特征多项式的根，其重数d_i（又叫做λ_i的代数重数）等于$E(\lambda_i)$（对应于λ_i的所有特征向量张成的特征子空间）的维数（又叫做λ_i的几何重数）。也就是说，T可对角化的一个充分必要条件是：每个特征值λ_i的代数重数d_i都等于其几何重数$\dim E(\lambda_i)$。或

$$d_i=\dim E(\lambda_i)=n-\mathrm{rank}(\lambda_i I-A)$$

$$i=1, 2, \cdots, n$$

因为一个对角阵的零度（其零空间的维数）等于其主对角元等于零的个数。而矩阵$m(T-\lambda_i I)_{\mathscr{B}}$，恰好有$d_i$个主对角元为零，故零度为$d_i$，而显然$E(\lambda_i)$的维数正好也是$d_i$，所以有如此的结论。这里只讲了$d_i=\dim E(\lambda_i)$ $i=1,2,\cdots,k$是T可对角化的必要条件，条件的充分性是显然的，請读者思考一下，为什么？下面的讨论也可解决这个问题。

若X是域F上的n维线性空间，T：$X\longrightarrow X$是一个线性变换，$f(\lambda)$是系数取自F的任一多项式，则据前面关于线性变换的线性运算的定义可知，$f(T)\in L(X,X)$。如果有$\lambda\in F$，$x\in X$满足$Tx=\lambda x$，则也有$f(T)x=f(\lambda)x$。由此我们可以证明如下的结论：

定理1.3.14 若线性变换T：$X\longrightarrow X$的所有不同的特征值为λ_1，λ_2，$\cdots$，$\lambda_k(k\leqslant n)$，$E(\lambda_1)$，$E(\lambda_2)$，$\cdots$，$E(\lambda_k)$是对应的特征子空间，且令$w_i=E(\lambda_i)i=1,2,\cdots,k$，$w=w_1+w_2+\cdots+w_k$。则有

$$w=w_1\oplus w_2\oplus\cdots\oplus w_k$$

或

$$\dim w=\sum_{i=1}^{k}\dim w_i$$

也就是说，w_1，w_2，$\cdots$，w_k是一组线性无关的子空间。

证明 记法$w=w_1+w_2+\cdots+w_k$本身说明w是T的所有特征向量张成的子空间。

要证明w是诸w_i的直和，据定义只要证明w中的零向量有唯一的表示就够了。为此设

$$x_1+x_2+\cdots+x_k=0,\quad x_i\in w_i$$

现在来证明$x_i=0$，$i=1,2,\cdots,k$。由于$x_i\in w_i$，故$Tx_i=\lambda_i x_i$，从而有$f(T)x_i=f(\lambda_i)x_i$，其中$f(\lambda)$为域F上的任一多项式。因此便得到

$$\begin{aligned}0=f(T)0&=f(T)x_1+f(T)x_2+\cdots+f(T)x_k\\&=f(\lambda_1)x_1+f(\lambda_2)x_2+\cdots+f(\lambda_k)x_k\end{aligned}$$

由于$f(\lambda)$是任取的，故可选取这样的多项式：

$$f_i(\lambda)=\prod_{\substack{j=1\\j\neq i}}^{k}(\lambda-\lambda_j)\Big/\prod_{\substack{l=1\\l\neq i}}^{k}(\lambda_i-\lambda_l)$$

$$i=1,\ 2,\ \cdots,\ k$$

很显然$f_i(\lambda)$有如下性质：

$$f_i(\lambda_j)=\delta_{ij},\ \text{即} f_i(\lambda_i)=1,\ f_i(\lambda_j)=0 \quad i\neq j$$

分别将$f_1(\lambda)$，$f_2(\lambda)$，…，$f_k(\lambda)$代入前式，便有

$$0=f_i(T)0=\sum_{j=1}^{k}f_i(T)x_j=\sum_{j=1}^{k}f_i(\lambda_j)x_j$$

$$=\sum_{j=1}^{k}\delta_{ij}x_j=x_i,\ i=1,\ 2,\ \cdots,\ k$$

这样便证明了

$$w=\bigoplus_{i=1}^{k}w_i$$

这个定理告诉我们：T关于不同特征值的特征向量是线性无关的。若B_i是T关于λ_i的一组线性无关的特征向量，则$\{B_1,B_2,\cdots,B_k\}$也是线性无关的。把前面讨论所得出的结果用定理的形式给出，则有：

定理1.3.15 设$T: X\longrightarrow X$是一个线性变换，λ_1，λ_2，…，λ_k为T的所有互异特征值，$w_i=E(\lambda_i)$，$\dim X=n$。则下述命题等价

(1) T可对角化

(2) T的特征多项式为$f(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_k)^{d_k}$且$\dim w_i=d_i$，$i=1,\ 2,\ \cdots,\ k$

(3) $\dim X=\dim w_1+\dim w_2+\cdots+\dim w_k$

例3. R^3上的线性变换T关于准标基底$\mathscr{B}=\{e_1,e_2,e_3\}$的矩阵表示为

$$m(T)_{\mathscr{B}} = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -1 \end{pmatrix} = A$$

则 $f(\lambda) = \det(\lambda I - A) = (\lambda - 1)(\lambda - 2)^2$，$\lambda_1 = 1$，$\lambda_2 = 2$，由于$\text{rank}(A - I) = 2$，$\text{rank}(A - 2I) = 1$，故有$\dim E(1) = 1 = d_1$，$\dim E(\lambda_2) = 2 = d_2$，所以能得出$T$可对角化的结论，即 $A \sim D = \text{diag}(1, 2, 2)$。进一步解方程组$(A - I)x = 0$及$(A - 2I)x = 0$可得

$$E(\lambda_1) = \text{span}(x_1 = (3, -1, 3)^T),$$

$$E(\lambda_2) = \text{span}(x_2 = (2, 1, 0)^T, x_3 = (2, 0, 1)^T)$$

若取 $\mathscr{B}' = \{x_1, x_2, x_3\}$，则便有$m(T)_{\mathscr{B}'} = D$，并且 $\mathscr{B} \xrightarrow{P} \mathscr{B}'$的基过渡阵$P = (x_1, x_2, x_3)$，从而

$$P^{-1}AP = D$$

从以上的讨论来看，要判断T或A可否对角化，都需要精确地计算出特征值λ_i，这是一个高次方程$f(\lambda) = 0$的求根问题。除非 $f(\lambda)$ 很特殊或次数很低，一般是求不出精确根的，而近似根λ_i代入方阵$(\lambda_i I - A)$，可能会使它的秩产生很大的变化。所以理论上很完满，实际上按这种方法去做不一定能够成功。我们再从另外一个角度，即用最小多项式的术语来分析T的可对角化的问题。

给定$T \in L(X, X)$，则对任意的自然数k有

$$I, T, T^2, \cdots, T^k \in L(X, X)$$

但由于$L(X, X) \cong F^{n \times n}$,故$\dim L(X, X) = n^2$。所以至多$k$取到$n^2$，$\{I, T, T^2, \cdots, T^k\}$便是线性相关的，从而存在一组标量$\alpha_0, \alpha_1, \cdots, \alpha_k \in F$使得

$$\alpha_0 I + \alpha_1 T + \cdots + \alpha_k T^k = 0$$

这说明多项式 $f(\lambda) = \sum_{i=0}^{k} \alpha_i \lambda^i$ 中的λ用T代替后，多项式$f(T)$

是一个零变换，即$f(T)=0$。我们把这样的多项式 $f(\lambda)$ 叫做T的零化多项式。显然每个T的零化多项式都是存在的，且有无穷多。第二章我们将要证明的 Cayley-Hamilton 定理，说明了T的特征多项式就是它的一个零化多项式。我们把T的所有零化多项式中次数最低的，首项系数为 1 的定义为T的最小多项式，记之为

$$m_T(\lambda)$$

明白点讲，多项式$m_T(\lambda)$应满足：

(1) $m_T(\lambda)$的最高次项的系数等于 1；

(2) $m_T(\lambda)$是T的一个零化多项式：$m_T(T)=0$；

(3) 如果$\varphi(\lambda)$是T的零化多项式，必有

$$m_T(\lambda)|\varphi(\lambda)$$

即$m_T(\lambda)$能整除T的一切零化多项式。很显然，T的最小多项式$m_T(\lambda)$是唯一的。若T有矩阵表示 $m(T)_{\mathscr{B}}=A$，则$m(f(T))_{\mathscr{B}}=f(A)$。故

$$f(T)=0 <\!-\!> f(A)=0$$

这表明T与A有相同的零化多项式和最小多项式，当然这里也考虑到了相似矩阵有相同的零化多项式，因为对任意的多项式$f(\lambda)$都有

$$f(P^{-1}AP)=P^{-1}f(A)P$$

这样，研究A与研究T一样有相同的结论。

据Cayley-Hamilton定理，T的特征多项式是T的零化多项式，故有$m_T(\lambda)|\det(\lambda I-A)$，这表明最小多项式$m_T(\lambda)$的根不能超出$T$的特征值。现在我们还可以证明：

定理1.3.16 T的特征多项式$f(\lambda)$与最小多项式$m_T(\lambda)$有相同的根（不计根的重数）。

证明 为此，只要证明，对$\lambda_0\in F$有

$$m_T(\lambda_0)=0 <\!-\!> f(\lambda_0)=0$$

就够了。

必要性。若$m_T(\lambda_0)=0$，则必有$m_T(\lambda)=(\lambda-\lambda_0)q(\lambda)$，且$q(\lambda)$的次数小于$m_T(\lambda)$的次数。据$m_T(\lambda)$的最小性，$q(T)\neq 0$。从而可取$x\in X$使$y=q(T)x\neq 0$，因此便有

$$0=m_T(T)x=(T-\lambda_0 I)q(T)x=(T-\lambda_0 I)y$$

这表明$0\neq y\in X$是T的关于特征值λ_0的特征向量，当然特征值λ_0是特征多项式$f(\lambda)$的根，即

$$f(\lambda_0)=0$$

充分性。若$f(\lambda_0)=0$，则λ_0是T的特征值，故存在$x\neq 0$使得$Tx=\lambda_0 x$，从而对任意的多项式$g(\lambda)$有$g(T)x=g(\lambda_0)x$，当然也使得

$$m_T(T)x=m_T(\lambda_0)x$$

但由于$m_T(T)=0$，故$m_T(\lambda_0)x=0$，又$x\neq 0$，从而有$m_T(\lambda_0)=0$。即T的任一特征值λ_0都是最小多项式$m_T(\lambda)$的根。

定理1.3.17 若T是可对角化的，λ_1，λ_2，…，λ_k是T的所有互异特征值，则必有

$$m_T(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_k)$$

证明 任取T的一个特征向量$x\in X$，则它至少满足下面k个方程中的一个：

$$(\lambda_i I-T)x=0,\ i=1,\ 2,\ \cdots,\ k$$

从而便有

$$(\lambda_1 I-T)(\lambda_2 I-T)\cdots(\lambda_k I-T)x=0$$

现令$p(T)=(\lambda_1 I-T)(\lambda_2 I-T)\cdots(\lambda_k I-T)$ 上式说明T的任一特征向量x都落在$p(T)$的零空间之中。由于T可对角化，即它有n个线性无关的特征向量$\{x_1, x_2, \cdots, x_n\}$可作为空间$X$的一个基底，任取$y\in X$，则有

$$y=\alpha_1 x_1+\alpha_2 x_2+\cdots+\alpha_n x_n$$

从而可知

$$p(T)y=\alpha_1 p(T)x_1+\alpha_2 p(T)x_2+\cdots+\alpha_n p(T)x_n$$

$$=0$$

这表明$p(T)$把整个X都变成了零向量，所以它是X上的一个零变换，即$p(T)=0$。这也说明多项式$p(\lambda)$是T的一个零化多项式。从而有

$$m_T(\lambda)\mid p(\lambda)$$

但$m_T(\lambda)$的次数不能再低于$p(\lambda)$的次数了，否则它将不再含有T的各种不同的特征值，这与前面的定理矛盾。故只有$m_T(\lambda)=p(\lambda)$。

这个定理指出了：可对角化的T，其最小多项式一定是不同的一次因式之积。其实，这个定理的逆也是成立的：若T的最小多项式是不同的一次因式之积，则T一定可对角化。不过，要证明它还需要其它概念。在证明定理1.3.22之后，便可解决。

例4. 求例1，2，3中的三个线性变换T的最小多项式。例3中的T是可对角化的，而特征多项式又是$f(\lambda)=(\lambda-1)(\lambda-2)^2$，所以很显然有

$$m_T(\lambda)=(\lambda-1)(\lambda-2)$$

例2中T的特征多项式也为$f(\lambda)=(\lambda-1)(\lambda-2)^2$，但它是不可对角化的，当然不能有

$$m_T(\lambda)=(\lambda-1)(\lambda-2)$$

但又知T的特征值为1，2，故只可能具有

$$m_T(\lambda)=(\lambda-1)^2(\lambda-2)\text{或}m_T(\lambda)=(\lambda-1)(\lambda-2)^2$$

再根据$m_T(\lambda)\mid f(\lambda)$，故只能是$m_T(\lambda)=(\lambda-1)(\lambda-2)^2$。要注意，决不能由此得出结论：不可对角化的$T$一定有$m_T(\lambda)=f(\lambda)$（$T$的特征多项式）。

例1中的T，其特征多项式为$f(\lambda)=\lambda^2+1$，不管域F取**R**还是取**C**，最小多项式都等于特征多项式。这是由于A不是纯量矩阵，$m_T(\lambda)$不能是一次的缘故。

虽然引入了最小多项式的概念，但对解决T可否对角化的问题仍然起不到决定性的作用，因为计算T（或A）的最小多项式

还没有有效方法。不过这个概念在线性代数中还是重要的。

对解决T可否对角化的问题，还有没有其它途径呢？现在我们再引入不变子空间的概念，以此来讨论T的最简矩阵表示。

定义1.3.18 设$T: X\longrightarrow X$是一个线性变换，$W\subset X$是X的一个线性子空间。若$TW\subset W$，即任取$x\in W$都有$Tx\in W$，则称W为T的不变子空间。

据定义，$\{0\}$，X，$\mathscr{R}(T)$，$\mathscr{N}(T)$，$E(\lambda_i)$都是T的不变子空间。任取T，$S\in L(X, X)$且$TS=ST$，则$\mathscr{R}(S)$，$\mathscr{N}(S)$皆是T的不变子空间。因为任取$x\in\mathscr{R}(S)$有$y\in X$使得$x=Sy$，从而有

$$Tx=TSy=STy\in\mathscr{R}(S)$$

任取$x\in\mathscr{N}(S)$有$Sx=0$，欲证$Tx\in\mathscr{N}(S)$，因为

$$S(Tx)=T(Sx)=T0=0$$

故$\mathscr{N}(S)$为T的不变子空间。这表明：凡是与T可交换的线性变换S，其值域与零空间皆是T的不变子空间。而T的任一多项式$p(T)$都与T可交换，故$p(T)$的值域和零空间也是T的不变子空间。特别是$(T-\lambda I)^k$ $(k=0, 1, \cdots)$的值域和零空间，其中包含了T的特征子空间，都是T的不变子空间。还有，不变子空间的和与交也是不变子空间。

当W是T的一个不变子空间时，我们将T的定义域限制在W上，便得到一个新的线性变换，叫做T在W上的限制，记为T_W，它除了和T的定义域有所不同外，其它并无区别。设不变子空间W的维数为r，$\{x_1, x_2, \cdots, x_r\}=\mathscr{B}'$为$W$的一个基底。把它扩充为

$$\mathscr{B}=\{x_1, x_2, \cdots, x_r, x_{r+1}, \cdots, x_n\}$$

使之成为X的一个基底，令$m(T)_{\mathscr{B}}=A$，则因为

$$Tx_1=a_{11}x_1+a_{21}x_2+\cdots+a_{r1}x_r+0x_{r+1}+\cdots+0x_n$$

$$Tx_2=a_{12}x_1+a_{22}x_2+\cdots+a_{r2}x_r+0x_{r+1}+\cdots+0x_n$$

$$\vdots$$

$$Tx_r = \alpha_{1r}x_1 + \alpha_{2r}x_2 + \cdots + \alpha_{rr}x_r + 0x_{r+1} + \cdots + 0x_n$$

$$Tx_{r+1} = \alpha_{1r+1}x_1 + \alpha_{2r+1}x_2 + \cdots$$
$$+ \alpha_{rr+1}x_r + \alpha_{r+1r+1}x_{r+1} + \cdots + \alpha_{nr+1}x_n$$

$$\vdots$$

$$Tx_n = \alpha_{1n}x_1 + \alpha_{2n}x_2 + \cdots + \alpha_{rn}x_r + \alpha_{r+1n}x_{r+1}$$
$$+ \cdots + \alpha_{nn}x_n$$

故有

$$A = \begin{bmatrix} C & D \\ 0 & E \end{bmatrix}, \quad m(T_W)_{\mathscr{B}'} = C$$

其中$C=(\alpha_{ij})1 \leqslant i, j \leqslant r$，$D=(\alpha_{ij})$，$1 \leqslant i \leqslant r$，$r+1 \leqslant j \leqslant n$，$E=(\alpha_{ij})$，$r+1 \leqslant i \leqslant n$，$r+1 \leqslant j \leqslant n$。这表明当$X=W \oplus W^c$且$W^c = \mathrm{span}(x_{r+1}, \cdots, x_n)$时，取$W$与$W^c$的基底合成$X$的基底$\mathscr{B}$，则$T$关于这个基底的矩阵表示$A$有较为简单的形式，其左下角一块为零。如果$W^c$也是$T$的不变子空间，显然$A$的右上角一块$D=0$，这时$A$呈块对角阵。由此可以想到：如果$w_1$，$w_2$，$\cdots$，$w_k$都是$T$的不变子空间，且$X = w_1 \oplus \cdots \oplus w_k$，则取诸$w_i$的基底合成$X$的基底$\mathscr{B}$，$T$关于$\mathscr{B}$的矩阵表示便是

$$A = \mathrm{diag}(A_1, A_2, \cdots, A_k)$$

A_i的阶数为w_i的维数。还可以进一步想到：若X能分解成T的一维不变子空间（由单个特征向量所张成）的直和，则T一定可对角化。由此可得出结论：

定理1.3.19 T可对角化的充要条件是X可以分解为T的一维不变子空间的直和。

根据上述分析，T的矩阵表示能有怎样简单的形式，取决于空间X能按T的不变子空间作怎样的直和分解。直和分解中不变子空间的个数越多，每个不变子空间的维数就越低，T的矩阵表示中每个小对角块的阶数也越小，从而越简单。至于每个小对角块的结构，又与对应的不变子空间的结构有关。如果要问：一个变换T

究竟有多少个不变子空间？它们的结构有什么特征？空间X能分解为其中的部分不变子空间的直和吗？对应这个直和分解，T的矩阵表示为何？这些问题要阐述清楚，需要做不少的准备，花很大的篇幅，所以这里只能这样给出一个结论：

如果X是复数域$\mathbf{C}$上的n维线性空间，T是X上的一个线性变换。则T的特征多项式$f(\lambda)$在$\mathbf{C}$上有n个根$\lambda_1, \lambda_2, \cdots, \lambda_n$，如果其中互异的共有$k$个：$\lambda_1, \lambda_2, \cdots, \lambda_k$，重数分别为$d_1, d_2, \cdots, d_k$，则

$$f(\lambda)=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_k)^{d_k}$$

据前面的讨论知T的最小多项式应为

$$m_T(\lambda)=(\lambda-\lambda_1)^{e_1}(\lambda-\lambda_2)^{e_2}\cdots(\lambda-\lambda_k)^{e_k}$$
$$1\leqslant e_i\leqslant d_i,\quad i=1, 2, \cdots, k$$

可以证明：若令：$w_i=\mathscr{N}((T-\lambda_i I)^{e_i})$，$i=1, 2, \cdots, k$，则易知每个$w_i$都是$T$的不变子空间，且

$$X=w_1\oplus w_2\oplus\cdots\oplus w_k$$

若B_i是w_i的基底，则$\mathscr{B}=\{B_1, B_2, \cdots, B_k\}$为$X$的一个基底。从而可知

$$m(T)_{\mathscr{B}}=J=\mathrm{diag}(J_1, J_2, \cdots, J_k)$$

其中$J_i=J_{m_i}(\lambda_i)=\mathrm{diag}(J_{i_1}(\lambda_i), J_{i_2}(\lambda_i), \cdots, J_{i_{l_i}}(\lambda_i))\in \mathbf{C}^{m_i\times m_i}, i=1, 2, \cdots, k$，$\sum_{i=1}^{k} m_i=n$。而$J_i$中的$J_{i_s}(\lambda_i)$有如下形式

$$J_{i_s}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \ddots & \ddots \\ & & & 1 \\ & & & \lambda_i\end{pmatrix}\in \mathbf{C}^{i_s\times i_s},\quad 1\leqslant i_s\leqslant m_i$$

并且还有$\sum\limits_{i=1}^{k}\sum\limits_{s=1}^{l_i} i_s=n$。$J_{i_s}(\lambda_i)$叫做一个 Jordan 块，而每个$J_i$都是若干个对应于特征值$\lambda_i$的 Jordan 块组成，$J$是由$k$个形如$J_i$的块组成，被称作为$T$的（或$T$的任一矩阵表示的）Jordan 标准形。也就是说，在复数域F上考虑，每个线性变换或每个复方阵都有一个 Jordan 标准形。说明白点就是：n维复线性空间上的线性变换都有一个 Jordan 形的矩阵表示；每个复方阵都与一个Jordan形的矩阵相似。还可证明，若不计较Jordan块的排列顺序，则每个方阵的 Jordan 标准形是唯一的。由于这个结论的证明需要太多的准备知识，且证明过程的本身实用性不大，所以暂略。但由于 Jordan 标准形有一系列良好的性质，在理论分析方面很有价值，所以在第二章末给出一种构造性的证明。

在结束T的最简矩阵表示的讨论之前，我们打算再完成两个结论的证明：

(1) T可对角化$\Longleftrightarrow T$的最小多项式$m_T(\lambda)$为不同一次因式的乘积。

(2) T可三角化$\Longleftrightarrow T$的最小多项式$m_T(\lambda)$为不同一次因式的幂积。

为此，我们引入传送子的概念，并证明一个引理。

定义1.3.20 设X是域F上的线性空间，$T: X\longrightarrow X$是一个线性变换，$W\subset X$是T的一个不变子空间，$F[\lambda]$表示系数取自F的所有多项式的集合，x是X中的任一向量，取$F[\lambda]$的子集：

$$S_T(x;W)=\{g(\lambda)\in F[\lambda]\mid g(T)x\in W\}$$

显然$S_T(x;W)\neq\phi$，因为T的特征多项式$f(\lambda)$、最小多项式以及所有零化多项式都属于这个集合。我们把$S_T(x;W)$中次数最低的首项系数等于 1 的多项式叫做$S_T(x;W)$的传送子。

据定义很容易证明：若$S_T(x;W)$的传送子为$g(\lambda)$，则必有$g(\lambda)\mid m_T(\lambda)$。证明如下：设

$$m_T(\lambda)=g(\lambda)p(\lambda)+r(\lambda)$$

$r(\lambda)$的次数低于$g(\lambda)$的次数且$r(\lambda)\neq 0$。则有

$$0=m_T(T)x=p(T)g(T)x+r(T)x$$

即$r(T)x=-p(T)g(T)x$。由于$g(T)x\in W$，而W为T的不变子空间，当然也是$p(T)$的不变子空间，故 $-p(T)g(T)x\in W$，从而$r(T)x\in W$，这与$g(\lambda)$的定义矛盾。从而说明 $r(\lambda)=0$，即有

$$g(\lambda)\mid m_T(\lambda)$$

引理1.3.21 设X是域F上的n维线性空间，而T：$X\longrightarrow X$是一个线性变换，且T的最小多项式$m_T(\lambda)$能分解为不同一次因式的幂积：

$$m_T(\lambda)=(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k}$$

又设$W\subset X$是T的真不变子空间（$W\neq X$）。则必存在 $x\in X$ 满足

(1) $x\notin W$（这表明$x\neq 0$），

(2) $(T-\lambda I)x\in W$，λ为T的某一特征值。

证明 由于$W\neq X$，故存在$y\in X$且$y\notin W$。令$g(\lambda)$是$S_T(y;W)$的传递子，则$g(\lambda)\mid m_T(\lambda)$。据$g(\lambda)$的定义知 $g(T)y\in W$，且据$y\notin W$知$g(\lambda)\neq$常数。而据$g(\lambda)\mid m_T(\lambda)$知

$$g(\lambda)=(\lambda-\lambda_1)^{e_1}(\lambda-\lambda_2)^{e_2}\cdots(\lambda-\lambda_k)^{e_k},$$

$$0\leqslant e_i\leqslant r_i$$

并且至少有一个$e_i\geqslant 1$。不妨设$e_1\geqslant 1$，则有

$$g(\lambda)=(\lambda-\lambda_1)h(\lambda)$$

据定义知$x=h(T)y\neq 0$且$x=h(T)y\notin W$，故有

$$(T-\lambda_1 I)x=(T-\lambda_1 I)h(T)y=g(T)y\in W$$

从而完成了证明。

定理1.3.22 X和T如引理所设，则有

$$T\text{可对角化}\Longleftrightarrow,\ m_T(\lambda)=\prod_{i=1}^{k}(\lambda-\lambda_i),\ \lambda_i\neq\lambda_j,\quad i\neq j。$$

证明 必要性。前面定理1.3.17已经证明。

充分性。令W是T的所有特征向量张成的子空间，它为T的不变子空间，且有

$$W = W_1 \oplus \cdots \oplus W_k, \quad W_i = E(\lambda_i), \quad i = 1, 2, \cdots, k$$

现在证明$W = X$。采取反证。若$\dim W < n$，则W为X的真子空间。据引理存在$x \notin W$且

$$(T - \lambda_i I)x \in W, \quad \lambda_i \text{为} T \text{的某一特征值}$$

令$y = (T - \lambda_i I)x \in W$，取$W_i$的基底$B_i$，则

$$\mathscr{B}' = \{ B_1, B_2, \cdots, B_k \}$$

为W的基底，或令$e_i = \dim W_i$，则有

$$\mathscr{B}' = \{ x_{11}, x_{12}, \cdots, x_{1e_1}, \cdots, x_{k1}, x_{k2}, \cdots, x_{ke_k} \}$$

由于$y \in W$，故有

$$y = \sum_{l=1}^{l_1} \alpha_{1l} x_{1l} + \sum_{l=1}^{l_2} \alpha_{2l} x_{2l} + \cdots + \sum_{l=1}^{l_k} \alpha_{kl} x_{kl}$$

任取T的一个多项式$h(T)$作用上式两端有

$$h(T)y = \sum_{l=1}^{e_1} \alpha_{1l} h(T) x_{1e} + \sum_{l=1}^{e_2} \alpha_{2l} h(T) x_{2l} + \cdots + \sum_{l=1}^{e_k} \alpha_{kl} h(T) x_{kl}$$

$$= \sum_{i=1}^{k} \sum_{l=1}^{e_i} \alpha_{il} h(\lambda_i) x_{il} \in W$$

现令$m_T(\lambda) = (\lambda - \lambda_i) q(\lambda)$，$q(\lambda)$为$m_T(\lambda)$的因式。

由于$(\lambda - a) | (\lambda^n - a^n)$，故有$(\lambda - a) | (f(\lambda) - f(a))$。据此有

$$(\lambda - \lambda_i) | (q(\lambda) - q(\lambda_i)),$$

$$q(\lambda) - q(\lambda_i) = (\lambda - \lambda_i) h(\lambda)$$

从而便有

$$\begin{aligned}(q(T) - q(\lambda_i) I)x &= q(T)x - q(\lambda_i)x \\ &= (T - \lambda_i I) h(T) x \\ &= h(T)(T - \lambda_i I)x\end{aligned}$$

$$=h(T)y$$

又由于 $h(T)y\in W$ 及 $0=m_T(T)x=(T-\lambda_i I)q(T)x$ 这表明 $q(T)x$ 是 T 的关于 λ_i 的特征向量或零向量，故有

$$q(T)x\in E(\lambda_i)=W_i\subset W$$

从而 $q(\lambda_i)x=q(T)x-h(T)y\in W$，但 $x\notin W$ 说明 $q(\lambda_i)=0$，否则 $q(\lambda_i)x\notin W$。这样就证明了 $(\lambda-\lambda_i)|q(\lambda)$，亦即 $(\lambda-\lambda_i)^2|m_T(\lambda)$，这与 $m_T(\lambda)$ 是不同一次因式的积相矛盾，故反设不真，应有 $W=X$。

定理1.3.23 X，T 如引理所设。则有

T 可三角化 $\Longleftrightarrow m_T(\lambda)=(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k}$

证明 必要性。若 T 可三角化，即

$$m(T)_{\mathscr{B}}=R=\begin{pmatrix} r_{11} & r_{12} & \cdots\cdots & r_{1n} \\ & r_{22} & \cdots\cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}$$

则 T 的特征多项式为

$$f(\lambda)=(\lambda-r_{11})(\lambda-r_{22})\cdots(\lambda-r_{nn})$$
$$=(\lambda-\lambda_1)^{d_1}(\lambda-\lambda_2)^{d_2}\cdots(\lambda-\lambda_k)^{d_k}$$

d_1，d_2，…，d_k 分别为不同对角元的重数。又根据 $m_T(\lambda)|f(\lambda)$ 及 $m_T(\lambda)$ 有与 $f(\lambda)$ 相同的根，故有

$$m_T(\lambda)=(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k}$$

$$1\leqslant r_k\leqslant d_k$$

充分性。若 $m_T(\lambda)=(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k}$，我们来构造 X 的一个基底 $\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$ 使得

$$m(T)_{\mathscr{B}}=R=\begin{pmatrix} r_{11} & r_{12} & \cdots\cdots & r_{1n} \\ & r_{22} & \cdots\cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix}$$

首先令$W_0=\{0\}$，它显然是T的真不变子空间，据引理1.3.21存在$x_1\in X$且$x_1\notin W_0$，使得

$$(T-r_{11}I)x_1\in W_0$$

此即$Tx_1=r_{11}x_1$，这表明r_{11}是T的特征值，x_1为相应的特征向量。再令$W_1=\text{span}(x_1)$，它也是T的真不变子空间，故存在$x_2\in X$且$x_2\notin W_1$满足$(T-r_{22}I)x_2\in W_1$，故可设$(T-r_{22}I)x_2=r_{12}x_1$，亦即$Tx_2=r_{12}x_1+r_{22}x_2$。令$W_2=\text{span}(x_1, x_2)$由于$W_2\neq X$且$TW_2\subset W_2$，故存在$x_3\in X$且$x_3\notin W_2$满足$(T-r_{33}I)x_3\in W_2$，不妨设

$$(T-r_{33}I)x_3=r_{13}x_1+r_{23}x_2$$

亦即
$$Tx_3=r_{13}x_1+r_{23}x_2+r_{33}x_3$$

依此类推，做到$W_n=\text{span}(x_1, x_2, \cdots, x_n)=X$为止，便得到$X$的基底$\mathscr{B}=\{x_1, x_2, \cdots, x_n\}$，且使

$$m(T)_{\mathscr{B}}=R=(r_{ij}),\ r_{ij}=0,\ i>j$$

推论1.3.24 若域$F=\mathbf{C}$，则任一$A\in\mathbf{C}^{n\times n}$都相似于一个上三角阵。

在第二章将证明一个更强的结论，即著名的舒尔引理。

五、正交变换、酉变换及其矩阵表示

n维内积空间是三维几何空间最直接的推广，它有类同于三维欧氏空间的几何结构，按照几何学的抽象定义：欧氏几何学是研究正交变换群下的不变量。所以正交变换在欧氏几何中占有很重要的位置。n维内积空间中的正交变换，在欧氏空间中叫正

交变换，在酉空间中叫做酉变换，象通常三维几何空间中的坐标旋转及镜象反射变换一样，能保持向量的长度和向量间的夹角不变，甚至一些几何对象都不发生形变。另外，这两种变换在理论分析和数值计算中也特别有用。

定义1.3.25 欧氏空间（酉空间）X上的线性变换T叫做正交（酉）变换，当它保持X中的任何两个向量的内积不变，即任取x，$y\in X$都有

$$(Tx|Ty)=(x|y)$$

并且把正交（酉）变换T在标准正交基下的矩阵表示叫做正交（酉）矩阵。

正交（酉）变换有几种等价的定义：

定理1.3.26 设T是欧氏（酉）空间X上的线性变换，其为正交（酉）变换的充分必要条件是下列之一：

(1) T保持向量的范数不变，也就是对任一$x\in X$，都有$\|Tx\|=\|x\|$；

(2) T把X的一个标准正交基底还映为一个标准正交基底，即当$\mathscr{B}=\{e_1,e_2,\cdots,e_n\}$且$(e_i|e_j)=\delta_{ij}$时，有$(Te_i|Te_j)=\delta_{ij}$；

(3) T关于任一标准正交基底$\mathscr{B}$的矩阵表示$m(T)_{\mathscr{B}}=Q(U)$满足$Q^T=Q^{-1}$（$U^H=U^{-1}$）；

(4) Q（U）的n个列向量构成$\mathbf{R}^n(\mathbf{C}^n)$的一个标准正交基底。

证明 (1) 若T是正交（酉）变换，据定义有

$$(Tx|Tx)=(x|x),\ \forall\ x\in X$$

即 $$\|Tx\|^2=\|x\|^2,\ \|Tx\|=\|x\|,$$

反之，若T保持范数不变，当然范数的平方亦不变，故对任意的x，$y\in X$有

$$(Tx|Tx)=(x|x),\ (Ty|Ty)=(y|y)$$

$$(T(x+y)|T(x+y))=(x+y|x+y)$$

利用内积的公理展开后面一个等式并消去相等的量便得到

$$(Tx|Ty)=(x|y)$$

注意，对酉变换T要使用内积的共轭对称性，先证明$\mathrm{Re}(Tx|Ty)=(x|y)$，再证明

$$\mathrm{Im}(Tx|Ty)=\mathrm{im}(x|y)$$

(2) 设$\{e_1, e_2, \cdots, e_n\}$是标准正交基底，$x, y\in X$ 则有$(e_i|e_j)=\delta_{ij}$，$x=\sum\xi_i e_i$，$y=\sum\eta_j e_j$。若T是正交（酉）变换，则有

$$(Te_i|Te_j)=(e_i|e_j)=\delta_{ij}$$

故$\{Te_1, Te_2, \cdots, Te_n\}$仍是标准正交基。

反之，若$(Te_i|Te_j)=\delta_{ij}$，则有

$$\begin{aligned}(x|y)&=(\sum\xi_i e_i|\sum\eta_j e_j)\\&=\sum_i\sum_j\xi_i\bar{\eta}_j(e_i|e_j)=\sum_i\sum_j\xi_i\bar{\eta}_j\delta_{ij}\\&=\sum_{i=1}^n\xi_i\bar{\eta}_i\end{aligned}$$

$$\begin{aligned}(Tx|Ty)&=(\sum\xi_i Te_i|\sum\eta_j Te_j)\\&=\sum_i\sum_j\xi_i\bar{\eta}_j(Te_i|Te_j)=\sum_i\sum_j\xi_i\bar{\eta}_j\delta_{ij}\\&=\sum_{j=1}^n\xi_i\bar{\eta}_i\end{aligned}$$

故有$(Tx|Ty)=(x|y)$。注意，当X是欧氏空间时，上述等式中$\bar{\eta}_j=\eta_j$。

(3) 设$\{e_1, e_2, \cdots, e_n\}=\mathscr{B}$是$X$的标准正交基底，则$(e_i|e_j)=\delta_{ij}$且令$m(T)_{\mathscr{B}}=Q=(q_{ij})$（或$U=(q_{ij})$）这时有

$$Te_i=\sum_{h=1}^n q_{hi}e_h$$

若T是正交（酉）变换，则据（2）知

$$(Te_i|Te_j)=\delta_{ij}$$

从而有

$$\begin{aligned}\delta_{ij}=(Te_i|Te_j)&=(\sum_k q_{ki}e_k|\sum_l q_{lj}e_l)\\&=\sum_k\sum_l q_{ki}\bar{q}_{lj}\delta_{kl}\\&=\sum_k q_{ki}\bar{q}_{kj}\end{aligned}$$

将它代入矩阵则有

$$(Q^TQ)_{ij}=\delta_{ij}，即Q^TQ=I$$

$$((U^HU)_{ij}=\delta_{ij}，即U^HU=I)$$

反之，从$Q^TQ=I(U^HU=I)$立即可得

$$\begin{aligned}(Q^TQ)_{ij}&=\sum_k q_{ki}\bar{q}_{kj}=\sum_k\sum_l q_{ki}\bar{q}_{lj}\delta_{kl}\\&=\sum_k\sum_l q_{ki}\bar{q}_{lj}(e_k|e_l)\\&=(\sum_k q_{ki}e_k|\sum_l q_{lj}e_l)\\&=(Te_i|Te_j)=\delta_{ij}\end{aligned}$$

说明T把标准正交基底仍映为标准正交基底。由（2）知 T 是正交（酉）变换。

（4） 若记$Q=(q_1q_2\cdots q_n)(=U)$，则有

$$Q^TQ=((q_j|q_i))=U^HU$$

$$(Q^TQ)_{ij}=(q_j|q_i)=(U^HU)_{ij}$$

根据（3）的证明有

$$(q_j|q_i)=\delta_{ij}$$

从而说明$\{q_1, q_2, \cdots, q_n\}$为$\mathbf{R}^n(\mathbf{C}^n)$的标准正交基底。反之亦然。

以下结论是显而易见的：

推论1.3.27 （1） 正交变换不改变向量间的夹角；

（2） 正交矩阵的行列式等于±1；

（3） 酉矩阵的行列式的模等于1；

（4） 正交矩阵的乘积、逆仍是正交矩阵；

(5) 酉矩阵之积、逆仍是酉矩阵；

(6) 正交阵与酉阵的特征值都落在单位圆周上。

第四节 例题与习题

1. 考虑线性空间$\mathbf{C}^3$中的下列子集：

(a) $\{x=(\xi_1, \xi_2, \xi_3)^T\in\mathbf{C}^3 | \xi_1\in\mathbf{R}\}$

(b) $\{x=(\xi_1, \xi_2, \xi_3)^T\in\mathbf{C}^3 | \xi_1=0\}$

(c) $\{x=(\xi_1, \xi_2, \xi_3)^T\in\mathbf{C}^3 |$ 或$\xi_1=0$或$\xi_2=0\}$

(d) $\{x=(\xi_1, \xi_2, \xi_3)^T\in\mathbf{C}^3 | \xi_1+\xi_2=0\}$

(e) $\{x=(\xi_1, \xi_2, \xi_3)^T\in\mathbf{C}^3 | \xi_1+\xi_2=1\}$

试判定其中哪些子集为$\mathbf{C}^3$的子空间。

2. 设$R[x]$为所有实系数多项式构成的$\mathbf{R}$上的线性空间，试判定下列子集中的哪些为$R[x]$的子空间：

(a) $\{p(x)\in R[x] | p(x)$的次数为5$\}$

(b) $\{p(x)\in R[x] | 2p(0)=p(1)\}$

(c) $\{p(x)\in R[x] | p(x)\geqslant 0, x\in[0, 1]\}$

(d) $\{p(x)\in R[x] | p(x)=p(1-x), \forall x\in\mathbf{R}\}$

3. 在线性空间$\mathbf{C}^n$中给定n的个向量

$$x_i=(a_{i1}, a_{i2}, \cdots, a_{in}), \ i=1, 2, \cdots, n$$

试证：$\{x_1, x_2, \cdots, x_n\}$线性相关的充分必要条件是$\det(A)=0$，其中

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

4. 设X是一个n维的线性空间。试证X中的任意$n+1$个向量都是线性相关的，而任意n个线性无关的向量都构成X的一个基底。

5. 如果把所有复数的集合 **C** 看作为实数域 **R** 上的一个线性空间，试问它的维数等于多少？

6. 如果把所有实数的集合 **R** 看作为有理数域 **Q** 上的一个线性空间，试问它的维数是多少？

7. 如果 Y，Z 都是线性空间 X 的子空间且维数相同，又知 $Y\subset Z$，试证 $Y=Z$。

8. 设 K，S 是多项式空间 $R[x]$ 中的两个子集，其定义是：

$$K=\{p(x)\in R[x]\mid p(-x)=-p(x),\ \forall x\in \mathbf{R}\}$$

$$S=\{p(x)\in R[x]\mid p(-x)=p(x),\ \forall x\in \mathbf{R}\}$$

证明

(a) K 与 S 都是 $R[x]$ 的子空间；

(b) K 与 S 互补，即 $R[x]=K\oplus S$。

9. 设 x，y，u，v 是 $\mathbf{C}^4$ 中的向量，且令 $M=\mathrm{span}(x, y)$，$N=\mathrm{span}(u, v)$，试问下列哪种情况使得 $\mathbf{C}^4=M\oplus N$ 为真？

(a) $x=(1, 1, 0, 0)^T$，$y=(1,0,1,0)^T$，$u=(0,1,0,1)^T$，$v=(0, 0, 1, 1)^T$。

(b) $x=(-1, 1, 1, 0)^T$，$y=(0, 1, -1, 1)^T$，$u=(1, 0, 0, 0)^T$，$v=(0, 0, 0, 1)^T$。

(c) $x=(1, 0, 0,1)^T$，$y=(0,1,1,0)^T$，$u=(1,0,1, 0)^T$，$v=(0, 1, 0, 1)^T$。

10. 设 X 是实数域 **R** 上的 n 维线性空间，$\{x_1, x_2,\cdots,x_n\}$ 为 X 的一个基底。任取 x，$y\in X$ 且有

$$x=\xi_1x_1+\xi_2x_2+\cdots+\xi_nx_n$$

$$y=\eta_1x_1+\eta_2x_2+\cdots+\eta_nx_n$$

我们把 x，y 的"内积"定义为

$$(x|y)=\xi_1\eta_1+2\xi_2\eta_2+\cdots+n\xi_n\eta_n$$

试证：$(x|y)$ 满足内积的四条公理。

11. 在线性空间 $\mathbf{R}^n$ 中，对任意向量

$$x=(\xi_1, \xi_2, \cdots, \xi_n)^T, \ y=(\eta_1, \eta_2, \cdots, \eta_n)^T$$

都可用任一给定的实对称正定阵$A\in\mathbf{R}^{n\times n}$，定义内积

$$(x|y)=y^TAx$$

试证：这个断语是正确的，并写出Cauchy-Schwarz不等式相对这种内积的具体形式。

12. 试证：在任意的欧氏空间E中，任取向量x，$y\in E$恒有

(a) $\|x+y\|^2+\|x-y\|^2=2\|x\|^2+2\|y\|^2$

(b) $(x|y)=\dfrac{1}{4}\|x+y\|^2-\dfrac{1}{4}\|x-y\|^2$

前者叫平行四边形公式，后者叫极化恒等式，其中$\|\cdot\|$是内积诱导的范数。

13. 设E是一个欧氏空间，$\dim E=n$，x_1，x_2，…，x_k，$y\in E$，且y可用$\{x_1, x_2, \cdots, x_k\}$线性表出。如果$y$与每个$x_i$（$i=1, 2, \cdots, k$）都正交，试证

$$y=0$$

14. 在酉空间$\mathbf{C}^n$中，如果两组基底

$$\{x_1, x_2, \cdots, x_n\}, \ \{y_1, y_2, \cdots, y_n\}$$

满足

$$(x_i|y_j)=\delta_{ij}, \qquad i, j=1, 2, \cdots, n$$

则称其中一个基底为另一个基底的伴随基底，有时也把它们叫做$\mathbf{C}^n$中的双正交族。试证：当给定基底$\{x_i\}_{i=1}^n$后，存在唯一的伴随基底$\{y_i\}_{i=1}^n$，并给出计算$\{y_i\}_{i=1}^n$的一种方法。

15. 利用欧氏空间$\mathbf{R}^n$中的Cauchy-Schwarz不等式证明：对任意的实数a_1，a_2，…，a_n，恒有不等式

$$\sum_{i=1}^n |a_i|\leqslant\sqrt{n(a_1^2+a_2^2+\cdots+a_n^2)}$$

成立。

16. 设$\{x_1, x_2, \cdots, x_m\}$是$n$维欧氏空间$E$中的一个标准正交向量组。试证：对任意向量$x\in E$都有不等式

$$\sum_{i=1}^{m}|(x|x_i)|^2 \leqslant \|x\|$$

成立。

17. 在n维多项式空间$R[x]_n$上定义变换如下：

$$T: R[\ ]_n \longrightarrow R[x]_n$$

$$p(x) \longmapsto p(x+1)$$

即$Tp(x)=p(x+1)$。并取$\mathscr{B}=\{1, x, \cdots, x^{n-1}\}$为$R[x]_n$的一个基底。

(a) 证明T是一个线性变换，

(b) 求出T关于基底$\mathscr{B}$的矩阵表示。

18. 给定$P=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$，现定义线性变换：

$$T: \mathbf{C}^{2\times 2} \longrightarrow \mathbf{C}^{2\times 2}$$

$$X \longmapsto PX$$

即$TX=PX$，并取$\mathbf{C}^{2\times 2}$的一个基底

$$\mathscr{B}=\left\{\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right\}$$

试求T关于基底$\mathscr{B}$的矩阵表示。

19. 设A，$B \in \mathbf{C}^{n\times n}$，且$AB=BA$。试证：

(a) 如果A有n个互异的特征值λ_1，λ_2，…，λ_n，则B一定与对角阵相似。

(b) 如果B与所有n阶方阵都可交换，则一定存在某一标量μ，使得$B=\mu I$。

20. 如果$\{x_1, x_2, \cdots, x_k\}$和$\{y_1, y_2, \cdots, y_k\}$是$n$维线性空间$X$中的两个线性无关组，则一定存在从$X$到$X$的一个可逆的线性变换$T$使得

$$Tx_i=y_i, \quad i=1, 2, \cdots, k$$

试证明之。

21. 如果线性变换T：$\mathbf{C}^2 \longrightarrow \mathbf{C}^2$关于$\mathbf{C}^2$的某一基底有矩阵表示$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$，试问，$T$有多少不变子空间。

22. 如果$A \sim B$，试证$A^n \sim B^n$，$A^T \sim B^T$。

23. 设A，$B \in \mathbf{C}^{n \times n}$，试证：如果$A$，$B$中有一个可逆，则$AB \sim BA$。

24. (a) 求一个三维空间上的线性变换T，使得T的值域为二维的。

(b) 求一个三维空间上的线性变换S，使得S的零空间为二维的。

25. 求一个矩阵$A \in \mathbf{C}^{4 \times 4}$，使得

$$\mathscr{R}(A) = \text{span}((1, 0, 1, 0)^T, (0, 1, 0, 1)^T)$$

26. 线性变换D：$R[x]_n \longrightarrow R[x]_n$

$$p(x) \longmapsto p'(x) = \frac{dp}{dx}$$

叫做微分算子。求D的秩及零度。

27. 给定A，$B \in \mathbf{C}^{n \times n}$且满足$AB = 0$，试证：

$$\text{rank}(A) + \text{rank}(B) \leqslant n$$

28. 试证：对每个$A \in \mathbf{C}^{n \times n}$，都存在$B \in \mathbf{C}^{n \times n}$使得

$$AB = 0 \text{且} \text{rank}(A) + \text{rank}(B) = n$$

29. 如果A，$B \in \mathbf{C}^{n \times n}$满足$A^2 = A$，$B^2 = B$。试证：

$$A \sim B \Longleftrightarrow \text{rank}(A) = \text{rank}(B)$$

30. 如果$A \in \mathbf{C}^{n \times n}$且$\text{rank}(A) = 1$。试证：存在唯一的标量$\alpha$使得$A^2 = \alpha A$。

31. 试证：任给一个n次的复系数多项式$f(\lambda)$，都可以找到一个方阵$A \in \mathbf{C}^{n \times n}$，使得$f(\lambda)$为$A$的特征多项式。

32. 如果$A \in \mathbf{C}^{n \times n}$的特征值为$\lambda_1$，$\lambda_2$，…，$\lambda_n$，试证：

$$\det(A)=\prod_{i=1}^{n}\lambda_i, \quad \operatorname{tr}(A)=\sum_{i=1}^{n}\lambda_i$$

33．试证：对任意的A，$B\in\mathbf{C}^{n\times n}$，$AB$与$BA$有相同的特征值。

［提示］$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}\begin{bmatrix} 0 & 0 \\ A & AB \end{bmatrix}\begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}=\begin{bmatrix} BA & 0 \\ A & 0 \end{bmatrix}$

34．试证：对任意的A，B，$C\in\mathbf{C}^{2\times 2}$，恒有

$$(AB-BA)^2C=C(AB-BA)^2$$

［提示］利用 Cayley-Hamilton 定理。

35．矩阵$A\in\mathbf{C}^{n\times n}$如果满足$A^k=I$（$k$是正整数），试问$A$是可对角化的吗？

36．设$A\in\mathbf{C}^{n\times n}$，$x\in\mathbf{C}^n$。若$A^{k-1}x\neq 0$，$A^kx=0$，则$\{x, Ax, \cdots, A^{k-1}x\}$是线性无关的。

37．（商空间，余维）设Y是线性空间X的一个子空间。所谓$x\in X$关于Y的一个陪集是指

$$x+Y=\{v|v=x+y,\ y\in Y\}$$

很显然，关于Y的所有陪集的集合

$$\{Z|Z=x+Y,\ x\in X\}$$

构成了X的一个分割。对这个集合定义代数运算：

$$(w+Y)+(x+Y)\hat{=}(w+x)+Y$$

$$\lambda(x+Y)\hat{=}\lambda x+Y$$

则构成一个线性空间，记为X/Y，叫做X关于Y的商空间。商空间X/Y的维数叫做Y的余维，即$\operatorname{codim}Y=\dim(X/Y)$。现设$X=\mathbf{R}^3$，$Y=\{(\xi_1,0,0)|\xi_1\in\mathbf{R}\}$，求$X/Y$，$X/X$，$X/\{0\}$并给出$X/Y$的几何表示。

38．（勾股定理）在欧氏空间E^n中，求证：

$$x \perp y \Longleftrightarrow \| x+y \|^2 = \| x \|^2 + \| y \|^2$$

而在酉空间（或更一般的复内积空间）中，有

$$x \perp y \Longrightarrow \| x+y \|^2 = \| x \|^2 + \| y \|^2$$

但反过来却未必成立，给出例子。

39． 若X是一个欧氏空间，证明：

$$\| x \| = \| y \| \Longrightarrow (x+y | x-y) = 0$$

如果$X=\mathbf{R}^2$，它的几何意义是什么？如果X是一个酉空间，又会得到怎样的结论？

40． 若X是一个内积空间，对一切$x \in X$都有

$$(x|u) = (x|v)$$

试证$u=v$。这个结论后面一章要用到。

41． 若X是一个n维的线性空间，$\{x_1, x_2, \cdots, x_n\}$是它的一个基底。试证：$X$上的一个内积完全由基向量的内积$\{\alpha_{ij} = (x_i | x_j)\}$确定。对于这样一组标量$\{\alpha_{ij}\}$能够以完全任意的方式选择吗？

42． 设X是一个内积空间。试证：

$x \perp y \Longleftrightarrow$对一切标量$\alpha$有$\| x+\alpha y \| = \| x-\alpha y \|$。

$x \perp y \Longleftrightarrow$对一切标量$\alpha$有$\| x+\alpha y \| \geqslant \| x \|$

如果$X=\mathbf{R}^2$，给出上述结论的几何解释。

43． （付立叶系数的极小性质）设X是一个内积空间，$\{e_1, e_2, \cdots, e_n\}$是$X$中的一组标准正交元。给定$x \in X$，令$y=\beta_1 e_1 + \beta_2 e_2 + \cdots + \beta_n e_n$，则$\| x-y \|$是$\beta_1, \beta_2, \cdots, \beta_n$的连续可微函数。试证明（用分析的方法即可）：$f(\beta_1, \beta_2, \cdots, \beta_n) = \| x-y \|$达到极小，当且仅当

$$\beta_i = (x | e_i), \quad i=1, 2, \cdots, n$$

44． 设$X=\mathbf{R}^4$，$a_1=(1, -1, 1, 1)^T$，$a_2=(5, 1, -3, 3)^T$，$x=(1, 3, -1, 3)^T$，$M=\mathrm{span}(a_1, a_2)$。试求$M$中与$x$

有最小夹角的矢量m_0。

45. 试求点$y=(1, 2, -1, 1)^T \in \mathbf{R}^4$到线性簇

$$V=\{v=x_0+\alpha_1 x_1+\alpha_2 x_2 \mid \alpha_1, \alpha_2 \in \mathbf{R}\}$$

的距离，其中$x_0=(0, -1, 1, 1)^T$，$x_1=(0, -3, -1, 5)^T$，$x_2=(4, -1, -3, 3)^T$。

提示：令$M=\mathrm{span}(x_1, x_2)$，则$V=x_0+M$，从而知点$y$到$V$的距离为

$$d(y, V)=\inf_{v \in V}\|y-v\|=\inf_{m \in M}\|y-x_0-m\|$$
$$=d(y-x_0, M)$$

这表明点y到V的距离等于点$y-x_0$到子空间M的距离。因此，只要求出$y-x_0$在M上的最佳逼近m_0，便求出了$d(y, V)=\|y-x_0-m_0\|$。

46. 在n维欧氏空间E^n中，给定m个线性无关的矢量$\{x_1, x_2, \cdots, x_m\}$，对它们作如下的线性组合（是一种很特殊的线性组合）：

$$\{\alpha_1 x_1+\alpha_2 x_2+\cdots+\alpha_m x_m \mid 0 \leqslant \alpha_i \leqslant 1,$$
$$i=1, 2, \cdots, m\}$$

这样所得到的矢量集合叫做以$\{x_1, x_2, \cdots, x_m\}$为顶点的平行多面体，记之为$V_m$。若以$\{x_1, x_2, \cdots, x_{m-1}\}$为顶点的平行多面体的体积为已知，仍记为$V_{m-1}$，而$x_m$到子空间$M=\mathrm{span}(x_1, \cdots, x_{m-1})$的距离$d$也已知，则$V_m$的体积(仍记为$V_m$)为：$V_m=V_{m-1}\times d$。试用归纳法证明这一结论。

47. 在欧氏空间$\mathbf{R}^4$中，求出两个单位长向量a_1, a_2，它们同时与下列向量

$$x=(2, 1, -4, 0)^T$$
$$y=(-1, -1, 2, 2)^T$$
$$z=(3, 2, 5, 4)^T$$

都正交。

48. 设 x_1，x_2，…，x_n，y 是欧氏空间E^m中的向量，且 y 可用x_1，x_2，…，x_n线性表出。证明：如果 y 与每个 x_i ($i=1$，2，…，n) 都正交，则$y=0$。

49. 设x_1，x_2，…，x_n是欧氏空间 E^m 中的一组向量，其Gram行列式定义为

$$g(x_1,x_2,\cdots,x_n)=\begin{vmatrix}(x_1|x_1) & (x_1|x_2) & \cdots & (x_1|x_n)\\(x_2|x_1) & (x_2|x_2) & \cdots & (x_2|x_n)\\ \vdots & \vdots & & \vdots\\(x_n|x_1) & (x_n|x_2) & \cdots & (x_n|x_n)\end{vmatrix}$$

试证：$\{x_1, x_2, \cdots, x_n\}$线性无关的充分必要条件是$g(x_1, x_2, \cdots, x_n)\neq 0$。

50. 设$\{e_1, e_2, \cdots, e_n\}$是 n 维欧氏空间 E^n 中的一个标准正交基底。又知

$$y_i=a_{i1}e_1+a_{i2}e_2+\cdots+a_{in}e_n$$
$$i=1,\ 2,\ \cdots,\ n$$

试证下面等式

$$g(y_1,\ y_2,\ \cdots,\ y_n)=\begin{vmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\a_{n1} & a_{n2} & \cdots & a_{nn}\end{vmatrix}$$

51. 设$\{e_1, e_2, e_3\}$是三维欧氏空间 E^3 中的一个标准正交基底。试求一个正交变换T：$E^3\longrightarrow E^3$，它使得：

$$T(e_1)=\frac{2}{3}e_1+\frac{2}{3}e_2-\frac{1}{3}e_3$$

$$T(e_2)=\frac{2}{3}e_1-\frac{1}{3}e_2+\frac{2}{3}e_3$$

52. 设$\{i, j, k\}$是$\mathbf{R}^3$中的一个标准正交基底。试求一个正交变换T：$\mathbf{R}^3\longrightarrow\mathbf{R}^3$，它使得$T(k)$为平面$x+y-1=0$的单位

法线矢量。

53. 证明：如果上三角阵

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ & a_{22} & a_{23} & \cdots & a_{2n} \\ & & a_{33} & \cdots & a_{3n} \\ & & & & \vdots \\ & & & & a_{nn} \end{pmatrix}$$

是正交矩阵，则A一定是对角阵，且主对角元a_{ii}（$i=1$，2，…，n）的绝对值$|a_{ii}|=1$。

54. 设一个3×3阶矩阵A的特征值为

$$\lambda_1=-2, \quad \lambda_2=1, \quad \lambda_3=3$$

相应的特征向量为

$$x_1=(1, 2, 2)^T, \quad x_2=(2, -2, 1)^T,$$
$$x_3=(-2, -1, 2)^T$$

试求出A。

55. 若线性变换T：$\mathbf{R}^4 \longrightarrow \mathbf{R}^4$关于$\mathbf{R}^4$的标准正交基底$\{e_1, e_2, e_3, e_4\}$的矩阵表示为

$$A=\begin{pmatrix} 5 & -1 & -2 & 0 \\ -1 & 5 & 0 & -2 \\ -2 & 0 & 5 & -1 \\ 0 & -2 & -1 & 5 \end{pmatrix}$$

现有正交矩阵

$$Q=\begin{pmatrix} 0.5 & -0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

把标准正交基底$\{e_1, e_2, e_3, e_4\}$变为$\{Qe_1, Qe_2, Qe_3, Qe_4\}$，试证$T$关于后一基底的矩阵表示为一对角阵，并求出$A$

的特征向量和特征值，也求出T的特征向量和特征值。并验证：对任意的向量x，$y\in\mathbf{R}^4$，都有

$$(Tx|y)=(x|Ty)$$

56．在使用矩阵处理问题时有一些很简单的结论需要记住并灵活运用，比如：

(1) 为了取出矩阵A的第i行，只要用e_i^T左乘A就行了，即e_i^TA为A的第i个行向量。同样，Ae_i^T为A的第i个列向量，而$e_i^TAe_j$为A的第(i,j)元素a_{ij}。

(2) 若$A\in\mathbf{C}^{m\times n}$，则有

对一切$x\in\mathbf{C}^n$有$Ax=0\Longleftrightarrow A=0$

(3) 若AXB有意义，则有

对一切X有$AXB=0\Longleftrightarrow A=0$或$B=0$

(4) 对一切x，y有$x^TAy=0\Longleftrightarrow A=0$

請读者按(1)证明(2)—(4)。

57．若X是一个n维的复线性空间，T：$X\longrightarrow X$是一个线性变换，$W\subset X$是T的一个不变子空间($TW\subset W$)，且$W\neq\{0\}$。试证：W中至少有一个T的特征向量。

58．若矩阵A，$B\in\mathbf{C}^{n\times n}$可交换，即$AB=BA$，则$A$的特征子空间必为$B$的不变子空间，从而可推出$A$与$B$有公共的特征向量。根据这个结论，证明两个可交换的对称方阵能够用一个非奇异阵P，同时化为对角阵，即$P^{-1}AP$与$P^{-1}BP$都是对角阵。注意使用“对称阵是可对角化”这一事实（在第二章第二节给出证明）。

59．对于$A\in\mathbf{C}^{n\times n}$的特征多项式

$$
\begin{aligned}
f(\lambda) &= \det(\lambda I - A) \\
&= \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} \\
&= \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n
\end{aligned}
$$

当n很大时，它的系数α_1，α_2，…，α_n是很难求出的。但可以证明有如下关系：

$$
\alpha_k = (-1)^k \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} \begin{vmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \cdots & a_{i_1 i_k} \\ a_{i_2 i_1} & a_{i_2 i_2} & \cdots & a_{i_2 i_k} \\ \vdots & \vdots & & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \cdots & a_{i_k i_k} \end{vmatrix}
$$

其中$k=1$，2，…，n。也就是α_k等于A的所有k阶主子式之和再乘上$(-1)^k$。特别是

$$
\alpha_1 = (-1)^1 \sum_{1 \leq i_1 \leq n} a_{i_1 i_1} = -\sum_{i=1}^{n} a_{ii} = -\operatorname{tr}(A)
$$

$$
\begin{aligned}
\alpha_n &= (-1)^n \sum_{1 \leq i_1 < i_2 < \cdots < i_n \leq n} A\begin{pmatrix} i_1, i_2, \cdots, i_n \\ i_1, i_2, \cdots, i_n \end{pmatrix} \\
&= (-1)^n \det(A)
\end{aligned}
$$

从而有

$$
\begin{aligned}
f(\lambda) &= \det(\lambda I - A) \\
&= \lambda^n - \lambda^{n-1}\operatorname{tr}(A) + \cdots \\
&\quad + \lambda^{n-k}(-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} A\begin{pmatrix} i_1, i_2, \cdots, i_k \\ i_1, i_2, \cdots, i_k \end{pmatrix} \\
&\quad + \cdots + (-1)^n \det(A)
\end{aligned}
$$

其中$A\begin{pmatrix} i_1, i_2, \cdots, i_k \\ i_1, i_2, \cdots, i_k \end{pmatrix}$表示由$A$的第$i_1$，$i_2$，…，$i_k$行和$i_1$，

i_2，…，i_k列交叉处元素所构成的k阶主子式。如果$f(\lambda)=0$的n个根为λ_1，λ_2，…，λ_n，则有

$$f(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$$

从而可得根与系数之间的关系式：

$$a_1=-\sum_{i=1}^{n}\lambda_i=-\mathrm{tr}(A)$$

$$a_2=\lambda_1\lambda_2+\lambda_1\lambda_3+\cdots+\lambda_{n-1}\lambda_n$$

$$=\sum_{1\leqslant i_1<i_2\leqslant n}A\binom{i_1,\ i_2}{i_1,\ i_2}$$

……………………………

$$a_k=(-1)^k\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}$$

$$=(-1)^k\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}A\binom{i_1,\ i_2,\cdots,\ i_k}{i_1,\ i_2,\cdots,\ i_k}$$

……………………………

$$a_n=(-1)^n\lambda_1\lambda_2\cdots\lambda_n$$

$$=(-1)^n\det(A)$$

特别是关系式

$$\sum_{i=1}^{n}\lambda_i=\mathrm{tr}(A),$$

$$\prod_{i=1}^{n}\lambda_i=\det(A)$$

要经常用到。請读者就$n=2$，3，4这些低阶的情况，利用行列式的求导法则及多项式的恒等来证明以上结论。

·60．证明方阵A可逆的充分必要条件是A没有等于零的特征值。

·61．证明实反对称阵A（$A^T = -A$）的所有特征值之和等于零。

第二章　矩阵的几种重要分解

寻求矩阵各种意义下的分解形式，无疑对与矩阵有关的数值计算和理论分析有着极为重要的意义。由于这些分解式的特殊形式一方面能明显地反映出原矩阵的某些数值特征，如矩阵的秩、行列式、特征值及奇异值等，另一方面分解的方法与过程往往提供了某些有效地的数值计算方法和理论分析根据。本章将介绍几种常用的矩阵分解：UR 分解（正交-三角分解）、R^TR 分解（Cholesky 分解）、谱分解、奇异值分解、秩分解和极分解等。此外，还将介绍在矩阵论中有重要应用的舒尔（Schur）引理，在此基础上给出正规矩阵的酉对角化问题，从而解决了实对称阵、实反对称阵、正交矩阵、厄尔米特阵、反厄尔米特阵、酉阵等的可对角化问题，它们是化二次型为标准形的理论基础。至于在计算方法中用到的矩阵的LU 分解，这里就省略了。

第一节　满秩方阵的UR分解及对称正定阵的$\mathbf{R}^T\mathbf{R}$分解

在第一章，我们通过 Gram-Schmidt 正交化过程，把线性内积空间的一般基底化为了标准正交基底。这个过程本身给出了又一个重要的结果，即所谓满秩方阵的UR分解。

定理2.1.1　设 $A\in\mathbf{C}_n^{n\times n}$，则存在酉矩阵$U$及正线（即主对角元皆为正的）上三角阵$R$，使得

$$A=UR$$

证明　将A按列分块写出，记之为

$$A=(a_1a_2\cdots a_n)$$

则$a_i\in \mathbf{C}^n$；由于$\mathrm{rank}(A)=n$，故$\{a_1, a_2, \cdots, a_n\}$构成酉空间$(\mathbf{C}^n, (\cdot|\cdot))$的一组基底。对它施行$G-S$正交化过程，得到一组标准正交基底：

$$\{u_1, u_2, \cdots, u_n\}$$

两组基底之间的换算关系为

$$u_1=a_1/\|a_1\|$$

$$u_k=\left(a_k-\sum_{i=1}^{k-1}(a_k|u_i)u_i\right)\Big/\left\|a_k-\sum_{i=1}^{k-1}(a_k|u_i)u_i\right\|$$

$$k=2, 3, \cdots, n$$

将上述关系式改写一下，便有

$$a_1=\|a_1\|u_1$$

$$a_k=\sum_{i=1}^{k-1}(a_k|u_i)u_i+\beta_k u_k$$

$$k=2, 3, \cdots, n$$

其中$\beta_k=\left\|a_k-\sum_{i=1}^{k-1}(a_k|u_i)u_i\right\|>0$，$k=1, 2, \cdots, n$。再写成矩阵的形式，便有

$$A=(a_1a_2\cdots a_n)=(u_1u_2\cdots u_n)\begin{pmatrix}\beta_1 & (a_2|u_1) & \cdots & (a_n|u_1)\\ 0 & \beta_2 & \cdots & (a_n|u_2)\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \beta_n\end{pmatrix}$$

若记$U=(u_1u_2\cdots u_n)$，上式右端的右因子为R，则U为一个酉阵，因为它是酉空间$(\mathbf{C}^n(\cdot|\cdot))$的一个标准正交基底；而$R$显然为正线上三角阵。

仿定理2.1.1的证明，有推论：

推论2.2.2 设$A\in \mathbf{R}_n^{n\times n}$，则存在正交阵$Q$及正线上三角阵$R$，使得$A=QR$。

满秩方阵的UR或QR分解有很多应用，兹举一例说明之。对于方程组$Ax=b$来说，如果$A\in \mathbf{R}_n^{n\times n}$，则有$A=QR$。在$Ax=$

b 两端同时左乘Q^T，则有 $Q^TAx=Q^Tb$。由于 $Q^TA=Q^TQR=R$，故得到等价的方程组$Rx=Q^Tb$。由于R是满秩的上三角阵，所以该方程组从最后一个方程求出最后一个未知数，代入倒数第二个方程求出倒数第二个未知数，再代入倒数第三个方程，依次向上回代，最后便求出所有未知数。又由于 Q^T 是一个正交矩阵，它左乘任一向量都不改变其范数，故可抑制计算过程中的误差积累。所以QR 分解在数值计算中是常用的工具之一。

对于列满秩的长方阵，有如下的分解式：

推论2.1.3 设$A\in \mathbf{C}_n^{m\times n}$，则存在酉矩阵$U\in\mathbf{C}^{m\times m}$，使得

$$A=UR$$

其中

$$R=\begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

这里 R_1为$n\times n$ 阶的正线上三角阵，0 为$(m-n)\times n$阶的零矩阵。

证明 由于A的秩为n，故它的n个列向量$\{a_1, a_2, \cdots, a_n\}$是线性无关的。在 $\mathbf{C}^m$ 中可适当选取向量$a_{n+1}, a_{n+2}, \cdots, a_m$，使$\{a_1, a_2, \cdots, a_m\}$为 $\mathbf{C}^m$ 的一个基底。对该基底施行Gram-Schmidt正交化，得到$\mathbf{C}^m$的一个标准正交基底$\{u_1,u_2,\cdots,u_m\}$。仿定理2.1.1可得

$$A=(a_1a_2\cdots a_n)=(u_1u_2\cdots u_m)\begin{pmatrix} \beta_1 & (a_2|u_1) & \cdots & (a_n|u_1) \\ 0 & \beta_2 & & (a_n|u_2) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \beta_n \\ 0 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

记上式最右端的左因子为U，右因子为R，便证明了所需要的结果。

考虑到$\{a_1, a_2, \cdots, a_n\}$在正交化的过程中是与$\{u_1,$

$u_2, \cdots, u_n\}$ 等价的，故有更为紧凑的形式：

$$A=(a_1a_2\cdots a_n)=(u_1u_2\cdots u_n)\begin{pmatrix}\beta_1 & (a_2|u_1) & \cdots & (a_n|u_1)\\ 0 & \beta_2 & \cdots & (a_n|u_2)\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \beta_n\end{pmatrix}$$

$$\triangleq UR$$

其中$U\in\mathbf{C}_n^{m\times n}$且满足$U^HU=I_n$，所以又叫做部分酉阵，也就是说它是酉矩阵的部分列向量构成的长方阵。而$R\in\mathbf{C}_n^{n\times n}$是一个正线上三角阵。以上两种分解形式叫做列满秩矩阵的UR分解，在数值分析中用起来是同样的方便。

对于一般的长方阵，有如下的正交分解：

定理2.1.4 设$A\in\mathbf{C}_r^{m\times n}$，则存在酉矩阵$U\in\mathbf{C}^{m\times m}$和$V\in\mathbf{C}^{n\times n}$及正线上三角阵$R\in\mathbf{C}_r^{r\times r}$，使得

$$A=U\begin{bmatrix}R & 0\\ 0 & 0\end{bmatrix}V^H$$

证明 由于A的秩为r，故A的n个列向量中有r个是线性无关的，其余的列向量可用之线性表出。我们通过列交换，也就是A右乘一个置换阵P_1(它是一个酉阵)，把这r个独立的列向量调到前r列的位置。我们把这个过程记为

$$AP_1=(y_1y_2\cdots y_ry_{r+1}\cdots y_n)$$

由于$\{y_{r+1}, \cdots, y_n\}$可用$\{y_1, y_2, \cdots, y_r\}$线性表出，故又有

$$AP_1=(y_1y_2\cdots y_r)[I_rC]$$

其中$C\in\mathbf{C}^{r\times(n-r)}$。又由于$(y_1y_2\cdots y_r)\in\mathbf{C}_r^{m\times r}$，据推论2.1.3有

$$(y_1y_2\cdots y_r)=U_1\begin{bmatrix}R_1\\ 0\end{bmatrix}$$

其中$U_1\in\mathbf{C}^{m\times m}$为酉矩阵，$R_1\in\mathbf{C}_r^{r\times r}$为正线上三角阵。从而有

$$AP_1=U_1\begin{bmatrix} R_1 \\ 0 \end{bmatrix}[I_r,C]=U_1\begin{bmatrix} R_1 & R_1C \\ 0 & 0 \end{bmatrix}$$

在上式中记 $B=[R_1 \quad R_1C]\in \mathbf{C}_r^{r\times n}$，则 $B^H\in \mathbf{C}_r^{n\times r}$。再据推论2.1.3，$B^H$ 有分解式：

$$B^H=V_1\begin{bmatrix} R_2 \\ 0 \end{bmatrix}$$

其中 $V_1\in \mathbf{C}^{n\times n}$ 为酉矩阵，$R_2\in \mathbf{C}_r^{r\times r}$ 为正线上三角阵，再将 $B=[R_2^H \quad 0]V_1^H$ 代入 AP_1 的上述分解式，便得到

$$AP_1=U_1\begin{bmatrix} R_2^H & 0 \\ 0 & 0 \end{bmatrix}V_1^H$$

由于 R_2^H 是正线下三角阵，故作适当的行及列的变换，可变成上三角阵，即在 R_2^H 的右、左侧分别乘上置换阵 P_2，$P_3\in \mathbf{C}^{r\times r}$，使 $R=P_2R_2^HP_3$ 为正线上三角阵。将 $R_2^H=P_2^TRP_3^T$ 代入已得到 AP_1 的分解式，并改写一下便得到

$$\begin{aligned} AP_1 &= U_1\begin{bmatrix} P_2^TRP_3^T & 0 \\ 0 & 0 \end{bmatrix}V_1^H \\ &= U_1\begin{bmatrix} P_2^T & 0 \\ 0 & I_{m-r} \end{bmatrix}\begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} P_3^T & 0 \\ 0 & I_{n-r} \end{bmatrix}V_1^H \end{aligned}$$

从而有

$$A=U_1\begin{bmatrix} P_2^T & 0 \\ 0 & I_{m-r} \end{bmatrix}\begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} P_3^T & 0 \\ 0 & I_{n-r} \end{bmatrix}V_1^HP_1^T$$

在上式中，记

$$U=U_1\begin{bmatrix} P_2^T & 0 \\ 0 & I_{m-r} \end{bmatrix},\quad V^H=\begin{bmatrix} P_3^T & 0 \\ 0 & I_{n-r} \end{bmatrix}V_1^HP_1^T$$

显然 U 及 V 皆为酉矩阵。从而证明了定理2.1.4。

对于实对称正定阵，还有如下特殊的分解。

定理2.1.5 设 A 是一个实对称正定矩阵，则存在正线上三角阵 R，使得

$$A=R^TR$$

此即所说的Cholesky分解，或三角-三角分解。

证明 由于实对称正定矩阵与单位阵合同，故有满秩阵P，使得$A=P^TP$。再据推论2.1.2有$P=QR$，从而有$A=P^TP=R^TQ^TQR=R^TR$。

Cholesky分解在解线性方程组时经常采用。如果方程组$Ax=b$的系数矩阵A是对称正定的（很多最小二乘问题的法方程就满足这一条件），则它等价于$R^TRx=b$。令$y=Rx$，则有$R^Ty=b$。由于R^T是正线下三角阵，很容易解出y，再就$Rx=y$，容易解出x，并且这种解法在计算机上有较好的稳定性。

长方阵的正交分解式在求解方程时也是比较有用的。如果$A\in\mathbf{C}_r^{m\times n}$，$b\in\mathbf{C}^m$，$A$的正交分解式为$U\begin{bmatrix}R_r & 0\\ 0 & 0\end{bmatrix}V^H$，则方程$Ax=b$等价于

$$\begin{bmatrix}R_r & 0\\ 0 & 0\end{bmatrix}V^Hx=U^Hb$$

令$V^Hx=y=(y_1,\ y_2)^T$，$U^Hb=(d_1,\ d_2)^T$，其中$y_1\in\mathbf{C}^r$，$y_2\in\mathbf{C}^{n-r}$，$d_1\in\mathbf{C}^r$，$d_2\in\mathbf{C}^{m-r}$。则原方程又等价于

$$\begin{cases}R_ry_1=d_1\\ 0=d_2\end{cases}$$

显然，若$d_2\neq0$，则方程$Ax=b$是不相容的，其最小二乘解的通式为

$$x=Vy=V\begin{bmatrix}R_r^{-1}d_1\\ c\end{bmatrix},\quad \forall c\in\mathbf{C}^{n-r}$$

而最小范数最小二乘解为

$$x_0=V\begin{bmatrix}R_r^{-1}d_1\\ 0\end{bmatrix}$$

若$d_2=0$，则方程$Ax=b$是相容的，其通解为

$$x=Vy=V\begin{bmatrix}R_r^{-1}d_1\\c\end{bmatrix},\quad \forall c\in\mathbf{C}^{n-r}$$

第二节 舒尔引理与正规矩陣

我们知道，不是每个方阵都能与对角阵相似。但是，每个实的或复的方阵在复数域上考虑，都有一个与之相似的若当（Jordan）标准形，它是一个特殊的上（或下）三角阵。这说明每个方阵在复数域上考虑都能与一个三角矩阵相似。借助于它，特别是它的Jordan标准形，在矩阵分析与计算中会带来很多方便。我们现在的问题是：怎样的矩阵能酉相似于一个对角阵？而一般的方阵能否酉相似于一个三角阵？因为酉矩阵比一般的满秩阵有更大的优点，比如酉矩阵作用在一个向量上不改变其（欧氏）范数；酉矩阵的逆等于它的共轭转置；酉矩阵的列向量组构成酉空间 $\mathbf{C}^n$ 的一个标准正交基底；正交相似（即酉相似的特例）关系与合同关系重合，等等。所以在酉相似下考虑问题，可以讨论的更为深入。本节从舒尔（Schur）引理入手，展开这方面的讨论。

定理2.2.1（Schur引理） 设 $A\in\mathbf{C}^{n\times n}$，则存在酉阵$U$使得

$$U^HAU=T=(\gamma_{ij}),\quad \gamma_{ij}=0,\quad i>j$$

也就是说，每个复方阵A都可酉相似于一个上三角阵T，T的主对角元为A的特征值。

证明 我们对A的阶数n，采取归纳法证明。当$n=1$，A本身就是一个上三角阵，所以定理是成立的。现假定定理对$n-1$阶方阵是成立的，也就是说，每个$n-1$阶矩阵A_1都存在一个酉阵$U_1\in\mathbf{C}^{(n-1)\times(n-1)}$，使得$U_1^HA_1U_1=T_1$是一个$n-1$阶上三角阵。由于在复数域上考虑，所以$A$总有特征值$\lambda_1\in\mathbf{C}$。设$A$对应$\lambda_1$的特征向量为$u_1\in\mathbf{C}^n$，且满足$\|u_1\|=1$。令$u_1$张成的一维子

空间为$[u_1]$，则

$$\mathbf{C}^n=[u_1]\dotplus[u_1]^{\perp}$$

其中$[u_1]^{\perp}$为$[u_1]$的正交补。取$[u_1]^{\perp}$的标准正交基底为$\{u_2, u_3, \cdots, u_n\}$，从而$\{u_1, u_2, \cdots, u_n\}$构成$\mathbf{C}^n$的一个标准正交基底。令$U_2=(u_1u_2\cdots u_n)$，则$U_2$为一个酉阵。从而有

$$(e_1e_2\cdots e_n)=I_n=U_2^HU_2=(U_2^Hu_1\ \ U^Hu_2\cdots U^Hu_n)$$

故有$U_2^Hu_1=e_1$；这样便推出

$$\begin{aligned}U_2^HAU_2&=U_2^HA(u_1u_2\cdots u_n)\\&=(U_2^HAu_1\ \ U_2^HAu_2\cdots U_2^HAu_n)\\&=(\lambda_1U_2^Hu_1\ \ U_2^HAu_2\cdots U_2^HAu_n)\\&=(\lambda_1e_1\ \ U_2^HAu_2\cdots U_2^HAu_n)\\&=\begin{bmatrix}\lambda_1 & b^T\\ 0 & A_1\end{bmatrix}\end{aligned}$$

其中$b\in\mathbf{C}^{n-1}$，$0\in\mathbf{C}^{n-1}$，$A_1\in\mathbf{C}^{(n-1)\times(n-1)}$。根据归纳假设，对$A_1$有酉阵$U_1$使得$U_1^HA_1U_1=T_1$，故取酉阵

$$U_3=\begin{bmatrix}1 & 0^T\\ 0 & U_1\end{bmatrix}\in\mathbf{C}^{n\times n}$$

其中$0\in\mathbf{C}^{n-1}$为零向量。再取$U=U_2U_3$，显然U为酉阵，且易验证$U^HAU=T$为上三角阵。又由于$U^H=U^{-1}$，故$A\sim T$。而相似阵有相同的特征值，故T的主对角元为A的全部特征值。

舒尔引理还有更为简捷的证明：每个复方阵A都有一个Jordan标准形，即$A=PJP^{-1}$，J为其Jordan标准形，是一个特殊的上三角阵。而满秩方阵P有UR分解，故有

$$A=PJP^{-1}=URJR^{-1}U^H$$

令$T=RJR^{-1}$，它显然是一个上三角阵，从而也推出$U^HAU=T$。

从舒尔引理可推出一系列重要的结果，下面给出几个比较直

接的推论。

定义2.2.2 若矩阵$A\in \mathbf{C}^{n\times n}$满足$A^HA=AA^H$，则称$A$是一个正规矩阵（*normal matrix*）。

读者可以验证实对称阵（$A^T=A$，$A\in\mathbf{R}^{n\times n}$）、实反对称阵（$A^T=-A$，$A\in\mathbf{R}^{n\times n}$）、厄尔米特矩阵（$A^H=A$，$A\in\mathbf{C}^{n\times n}$）、反厄尔米特矩阵（$A^H=-A$，$A\in\mathbf{C}^{n\times n}$）、正交矩阵（$A^T=A^{-1}$，$A\in\mathbf{R}^{n\times n}$）及酉矩阵（$U^H=U^{-1}$，$U\in\mathbf{C}^{n\times n}$）都是正规矩阵。这说明正规矩阵包括很大一类常用的矩阵。关于正规矩阵有如下重要的性质：

定理2.2.3 $A\in\mathbf{C}^{n\times n}$，$A^HA=AA^H \Longleftrightarrow$存在酉阵$U$及对角阵$\Lambda$，使得$A=U\Lambda U^H$。也就是说，一个方阵酉相似于一个对角阵，当且仅当它是一个正规阵。

证明 必要性。若A是正规阵，即$A^HA=AA^H$。据舒尔引理存在酉阵U使得$A=UTU^H$，从而可推出有$TT^H=T^HT$。记

$$T=\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ & \gamma_{22} & \cdots & \gamma_{2n} \\ & & \ddots & \vdots \\ & & & \gamma_{nn} \end{pmatrix}$$

则$(TT^H)_{ii}=\sum_{j=i}^{n}|\gamma_{ij}|^2$，$(T^HT)_{ii}=\sum_{j=1}^{i}|\gamma_{ji}|^2$，因$(TT^H)_{ii}=(T^HT)_{ii}$，$i=1，2，\cdots，n$，故有$\gamma_{ij}=0$，$i\neq j$。这表明$T$是一个对角阵，记$\Lambda=\mathrm{diag}(\gamma_{11}，\gamma_{22}，\cdots，\gamma_{nn})$，则$A=U\Lambda U^H$。

充分性。若A酉相似于对角阵，即$A=U\Lambda U^H$，则显然有

$$\begin{aligned} A^HA &= U\Lambda^HU^HU\Lambda U^H=U\Lambda^H\Lambda U^H \\ &= U\Lambda\Lambda^HU^H=U\Lambda U^HU\Lambda^HU^H \\ &= AA^H \end{aligned}$$

这说明A是一个正规阵。

上面的定理表明：

推论2.2.4 $A\in\mathbf{C}^{n\times n}$是正规阵，当且仅当$A$有$n$个特征向量构成酉空间$\mathbf{C}^n$的一个标准正交基底。

推论2.2.5 $A\in\mathbf{R}^{n\times n}$是正规阵，$r_1$，$r_2$，…，$r_k$是$A$的实特征值，$c_t\pm id_t$（$t=1, 2, \cdots, p$，$2p+k=n$）是$A$的复特征值（因为$A$的特征多项式是实系数多项式，故复根共轭成对的出现）。则存在正交矩阵$U\in\mathbf{R}^{n\times n}$，使得

$$U^TAU=\mathrm{diag}(r_1, r_2, \cdots, r_k)\dot{+}\sum_{t=1}^{p}{}^{\cdot}\begin{pmatrix}c_t & -d_t\\ d_t & c_t\end{pmatrix}$$

(提示：若$x+iy$是A对应于特征值$c+id$的特征向量，则$x-iy$便是A关于$c-id$的特征向量,其中的x，$y\in\mathbf{R}^n$。由此可证$(x|y)=0$且$A(y\ x)=(y\ x)\begin{bmatrix}c & -d\\ d & c\end{bmatrix}$)。

推论2.2.6 若A是实对称矩阵，则A的特征值λ_1，λ_2，…，λ_n皆为实数，且存在正交阵$U\in\mathbf{R}^{n\times n}$，使得

$$U^TAU=U^{-1}AU=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

推论2.2.7 若$A\in\mathbf{C}^{n\times n}$有重特征值，则一定存在与$A$任意接近且有互异特征值的矩阵$B$，这里的$A$与$B$的接近可以认为是对应元素的接近。

推论2.2.8 $A\in\mathbf{C}^{n\times n}$，$A^r\to0$ $(r\to\infty)$ $\Longleftrightarrow|\lambda_i(A)|<1$，$i=1, 2, \cdots, n$；其中$A^r\to0$是指$A^r$的每个元素$(A^r)_{ij}\to0$；$\lambda_i(A)$表示$A$的第$i$个特征值。

证明 必要性。设$A^r\to0$。采用反证法。若A有某一个特征值λ满足$|\lambda|\geqslant1$，不妨设其对应的特征向量为$u\neq0$，从而有$Au=\lambda u$及$A^ru=\lambda^ru$。显然，当$r\to\infty$时有$\lambda^ru\not\to0$，故应有$A^r\not\to0$，这与原设矛盾。

充分性。设$|\lambda_i(A)|<1$，$i=1, 2, \cdots, n$；若λ_1，…，λ_n互异，则有满秩阵C，使得$A=C\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)C^{-1}$，

记$\Lambda=\text{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$，则有$A^r=C\Lambda^rC^{-1}$。显然，当$r\to\infty$时有$\Lambda^r\to0$，故$A^r\to0$。若$\lambda_1,\lambda_2,\cdots,\lambda_n$有相同者，据舒尔引理总有$A=UTU^H$，$A^r=UT^rU^H$。这表明$A^r\to0$ $\Longleftrightarrow T^r\to0$。设上三角阵$T$为

$$T=\begin{pmatrix}\lambda_1 & t_{12} & \cdots & t_{1n}\\ & \lambda_2 & \cdots & t_{2n}\\ & & \ddots & \vdots\\ & & & \lambda_n\end{pmatrix}$$

现构造上三角阵M如下：

$$M=\begin{pmatrix}\mu_1 & \alpha & \cdots & \alpha\\ & \mu_2 & \cdots & \alpha\\ & & \ddots & \vdots\\ & & & \mu_n\end{pmatrix}$$

其中$\alpha\geqslant\max\limits_{i,j}|t_{ij}|$，$\max\limits_{i}|\lambda_i|<\mu_1<\mu_2<\cdots<\mu_n<1$，则首先据前面的证明知$M^r\to0$，故$(M^r)_{ij}\to0$。另外，据$\alpha\geqslant\max\limits_{i,j}|t_{ij}|$可推出$|(T^r)_{ij}|\leqslant(M^r)_{ij}$，从而知$(T^r)_{ij}\to0$，这便证明了$T^r\to0$。从而推出了$A^r\to0$。

推论2.2.9 设$A\in\mathbf{C}^{n\times n}$的特征多项式为

$$\phi(\lambda)=\det(\lambda I-A)=\lambda^n+\alpha_1\lambda^{n-1}+\cdots+\alpha_{n-1}\lambda+\alpha_n$$

则一定有

$$\phi(A)=A^n+\alpha_1A^{n-1}+\cdots+\alpha_{n-1}A+\alpha_nI=0$$

这就是线性代数中有名的凯莱-哈密尔顿定理。

证明 若A的特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$互异，则有满秩阵$C$使得$A=C\Lambda C^{-1}$，其中$\Lambda=\text{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$。从而有$\phi(A)=C\phi(\Lambda)C^{-1}=C\text{diag}(\phi(\lambda_1),\phi(\lambda_2),\cdots,\phi(\lambda_n))C^{-1}$。由于每个$\phi(\lambda_i)=0$，故有$\phi(A)=0$，推论获证。若$A$有重特征值，则据推论2.2.7可以构造一个矩阵序列$\{B_m\}$，且满足以下两个条件：每个B_m都有互异的n个特征值；$B_m\to A$。现在记

$$\psi_m(\lambda)=\det(\lambda I-B_m)$$
$$=\lambda^n+\beta_1^{(m)}\lambda^{n-1}+\cdots+\beta_{n-1}^{(m)}\lambda+\beta_n^{(m)}$$

由于特征多项式的系数$\beta_i^{(m)}$是矩阵B_m的元素的多项式函数（行列式的展开说明了这一点），故是B_m诸元素的连续函数。因此，当$B_m\to A$时有$\beta_i^{(m)}\to\alpha_i$，故而有

$$\phi(A)=A^n+\alpha_1A^{n-1}+\alpha_2A^{n-2}+\cdots+\alpha_{n-1}A+\alpha_nI$$
$$=\lim_{B_m\to A}(B_m^n+\beta_1^{(m)}B_m^{n-1}+\beta_2^{(m)}B_m^{n-2}+\cdots$$
$$+\beta_{n-1}^{(m)}B_m+\beta_n^{(m)}I)$$
$$=\lim_{B_m\to A}\psi_m(B_m)$$

但是，由于每个B_m的n个特征值是互异的，故据前面的证明有$\psi_m(B_m)=0$，这便证明了$\phi(A)=0$。

凯莱-哈密尔顿定理有很多应用，也有很多的证明方法。如果把$\phi(A)=0$改写为

$$A^n=-(\alpha_1A^{n-1}+\alpha_2A^{n-2}+\cdots+\alpha_{n-1}A+\alpha_nI)$$

这表明A^n可用$\{I, A, A^2, \cdots, A^{n-1}\}$线性表出，同样可推出对每个正整数$m\geqslant n$，$A^m$也可用之线性表出。这个结果在后面的矩阵分析中会用到。如果A^{-1}存在，则据$\alpha_n=(-1)^n\det(A)\neq0$及$\phi(A)=0$可推出

$$A^{-1}=-\frac{1}{\alpha_n}(A^{n-1}+\alpha_1A^{n-2}+\cdots+\alpha_{n-1}I)$$

这既表明A^{-1}也是A的$n-1$次多项式，同时也给出了求A^{-1}的一种方法。

从上面的讨论可以看到舒尔引理在矩阵论中的重要作用。不仅如此，舒尔引理还给出了著名的舒尔不等式（推导过程见定理5.1.1）：

$$\sum_{i=1}^n|\lambda_i|^2\leqslant\mathrm{tr}(A^HA)=\sum_{i=1}^n\sum_{j=1}^n|a_{ij}|^2$$

其中 λ_i 为 $A=(a_{ij})\in\mathbf{C}^{n\times n}$ 的特征值。这个不等式之所以有名，倒不在于它给出了 A 的特征值的一个整体估计，而是它有多方面的应用。现在我们举一个它在矩阵的秩的估计方面的应用来说明。

任取 $M\in\mathbf{C}^{n\times n}$，$1\leqslant k\leqslant n-1$，将 M 分块为

$$M=\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

其中 A 是 M 的 k 阶首主子式，D 是 $n-k$ 阶主子式。这里设 $B\neq0$，$C\neq0$，否则 M 的秩或行列式都可降阶处理。并记

$$l_k=\mathrm{tr}(AA^H)+\mathrm{tr}(DD^H)+2\sqrt{\mathrm{tr}(BB^H)\mathrm{tr}(CC^H)}$$

$$l=\min_{1\leqslant k\leqslant n} l_k$$

则有如下结论：

定理2.2.10 M，l 如上所设，$\mathrm{rank}(M)$ 为 M 的秩，则有

$$\mathrm{rank}(M)\geqslant\frac{1}{l}|\mathrm{tr}(M)|^2$$

证明 令 λ_1，λ_2，…，λ_n 为 M 的特征值，由舒尔引理知，存在酉阵 U 使得

$$UMU^H=R=\begin{bmatrix} \lambda_1 & & * \\ & \lambda_2 & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

设 M 共有 s 个非零特征值，则 $s\leqslant\mathrm{rank}(M)$。不失一般性，令 λ_1，…，λ_s 为非零特征值，于是

$$|\mathrm{tr}(M)|^2=\left|\sum_{i=1}^{n}\lambda_i\right|^2=\left|\sum_{i=1}^{s}\lambda_i\right|^2$$

再利用Cauchy-Schwarz不等式，便有

$$|\mathrm{tr}(M)|^2=\left|\sum_{i=1}^{s}\lambda_i\right|^2\leqslant s\sum_{i=1}^{s}|\lambda_i|^2$$

$$\leqslant \operatorname{rank}(M)\sum_{i=1}^{n}|\lambda_i|^2$$

取 $\xi_1^2=\operatorname{tr}(BB^H)>0$，$\xi_2^2=\operatorname{tr}(CC^H)>0$，并构造方阵

$$K=\begin{pmatrix} A & \left(\dfrac{\xi_2}{\xi_1}\right)^{1/2}B \\ \left(\dfrac{\xi_1}{\xi_2}\right)^{1/2}C & D \end{pmatrix}$$

则有

$$KK^H=\begin{pmatrix} AA^H+\dfrac{\xi_2}{\xi_1}BB^H & * \\ * & \dfrac{\xi_1}{\xi_2}CC^H+DD^H \end{pmatrix}$$

从而有

$$\operatorname{tr}(KK^H)=\operatorname{tr}(AA^H)+\operatorname{tr}(DD^H)+2\xi_1\xi_2=l_k$$

又因为有等式

$$K=\begin{pmatrix} \left(\dfrac{\xi_2}{\xi_1}\right)^{1/2}I_k & 0 \\ 0 & I_{n-k} \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} \left(\dfrac{\xi_1}{\xi_2}\right)^{1/2}I_k & 0 \\ 0 & I_{n-k} \end{pmatrix}$$

故$K\sim M$，所以λ_1，λ_2，…，λ_n也是K的特征值。从而据舒尔不等式有

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant \operatorname{tr}(KK^H)$$

与前面推出的不等式相结合，便有

$$|\operatorname{tr}(M)|^2\leqslant \operatorname{rank}(M)\sum_{i=1}^{n}|\lambda_i|^2\leqslant \operatorname{rank}(M)\operatorname{tr}(KK^H)$$

$$=l_k\operatorname{rank}(M)$$

由于上面不等式对$k=1$，2，…，n均成立，故又有

$$|\operatorname{tr}(M)|^2\leqslant l\operatorname{rank}(M)$$

亦即

$$\mathrm{rank}(M) \geqslant \frac{1}{l}|\mathrm{tr}(M)|^2$$

这个不等式给出了 rank(*M*) 的一个下界。1966 年 Ky Fan 与 Hoffman曾给出了另一个方阵*A*的秩的下界估计式：

$$\mathrm{rank}(A) \geqslant \sum_{i=1}^{n} \frac{|a_{ii}|}{\sum_{j=1}^{n}|a_{ij}|}$$

当和式中出现$\frac{0}{0}$时，规定取 0 。如果取方阵

$$M=\begin{pmatrix} 3 & 3 & 1 \\ 1 & 4 & 2 \\ 1 & 2 & 4 \end{pmatrix}$$

则很容易算得

$$l_1=3^2+(4^2+2^2+2^2+4^2)+2\sqrt{(3^2+1^2)(1^2+1^2)}$$
$$=49+4\sqrt{5}$$
$$l_2=(3^2+3^2+1^2+4^2)+4^2+2\sqrt{(1^2+2^2)(1^2+2^2)}$$
$$=61$$
$$l=l_1=49+4\sqrt{5}, \ \mathrm{tr}(M)=11,$$

利用定理2.2.10中的估计式有

$$\mathrm{rank}(M) \geqslant \frac{|\mathrm{tr}(M)|^2}{l}=\frac{121}{49+4\sqrt{5}}>2$$

利用Ky Fan-Hoffman 的估计式有

$$\mathrm{rank}(M) \geqslant \frac{3}{7}+\frac{4}{7}+\frac{4}{7}=1+\frac{4}{7}=\frac{11}{7}$$

这表明对某些方阵，定理2.2.10中的估计式更精确些。

推论2.2.11 M，l 与定理2.2.10同。如果

$$|\mathrm{tr}(M)|^2>(n-1)l$$

则$\det(M)\neq 0$。因为这时有

$$\text{rank}(M)\geqslant\frac{1}{l}|\text{tr}(M)|^2>\frac{1}{l}(n-1)l=n-1$$

如果取

$$M=\begin{pmatrix} 5 & 0 & 1 & 0 & 1 & -1 \\ 1 & 5 & -1 & -1 & -1 & 0 \\ -1 & -1 & 4 & -1 & -1 & -1 \\ 0 & 1 & 0 & 5 & 1 & -1 \\ 2 & 0 & -2 & -1 & 5 & 1 \\ 0 & -1 & 0 & 1 & 0 & 5 \end{pmatrix}$$

直接计算知$l=l_4=150+2\sqrt{70}$，$\text{tr}(M)=29$，容易看出$|\text{tr}(M)|^2=29^2>(6-1)l$，即知$\text{rank}(M)>5$，说明$M$是非奇异的。但Ky Fan-Hoffman估计式只能给出$\text{rank}(M)>3$，判定不了M的非奇异性。

据定理2.2.10的推导，有下面两个不等式

$$|\text{tr}(M)|^2\leqslant\text{rank}(M)\sum_{i=1}^n|\lambda_i|^2,\quad \sum_{i=1}^n|\lambda_i|^2\leqslant\text{tr}(MM^H)$$

关于任意方阵M都成立。从而得到

$$\text{rank}(M)\geqslant|\text{tr}(M)|^2/\text{tr}(MM^H)$$

因此，对任意长方阵$A\in\mathbf{C}^{m\times n}$，取$M=AA^H$代入上面不等式，便有如下结论：

定理2.2.12 任取$A\in\mathbf{C}^{m\times n}$，则有

$$\text{rank}(A)\geqslant[\text{tr}(AA^H)]^2/\text{tr}((AA^H)^2)\quad m\leqslant n$$

$$\text{rank}(A)\geqslant[\text{tr}(A^HA)]^2/\text{tr}((A^HA)^2)\quad m>n$$

推论2.2.13 对$A\in\mathbf{C}^{m\times n}$，若有

$$[\text{tr}(A^HA)]^2>(n-1)\text{tr}[(A^HA)^2]$$

则A^HA是满秩的，从而是正定Hermite阵。

第三节　幂等矩阵、投影算子以及矩阵的谱分解式

前面已经证明过：若 $A\in\mathbf{C}^{n\times n}$ 有 n 个线性无关的特征向量：$\{x_1, x_2, \cdots, x_n\}$，令 $P=(x_1\ x_2\ \cdots\ x_n)\in\mathbf{C}_n^{n\times n}$，则有 $P^{-1}AP=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\triangleq\Lambda$，其中 λ_i 为对应于 x_i 的特征值。这叫做 A 是可对角化的。反之亦然。另外，$\{x_1, x_2, \cdots, x_n\}$ 也构成了 $\mathbf{C}^n$ 的一个基底。设 $S_i=[x_i]$ 是由 x_i 张成的一维子空间，并把 A：$\mathbf{C}^n\to\mathbf{C}^n$ 视为一个线性变换，则 S_i 是 A 的不变子空间，并且 $\mathbf{C}^n=\sum\limits_{i=1}^{n}{}' S_i$。对任意的 $x\in\mathbf{C}^n$，有唯一的分解式：

$$x=y_1+y_2+\cdots+y_n,\quad y_i\in S_i$$

若令 P_i 为沿着子空间 $\sum\limits_{\substack{j=1\\j\neq i}}^{n}{}' S_j$ 向 S_i 的投影（注意，不一定是正交投影），则有

$$P_i x=y_i,\quad i=1, 2, \cdots, n$$

从而有

$$x=\sum_{i=1}^{n} y_i=\sum_{i=1}^{n} P_i x=\left(\sum_{i=1}^{n} P_i\right)x$$

据 $x\in\mathbf{C}^n$ 的任意性，说明 $\sum\limits_{i=1}^{n} P_i=I$。另一方面，$x$ 又可用 $\{x_1, x_2, \cdots, x_n\}$ 线性表出：

$$x=\sum_{i=1}^{n}\alpha_i x_i,\quad \alpha_i\in\mathbf{C},\quad \alpha_i x_i\in S_i,\quad i=1, 2, \cdots, n$$

再考虑到 x 关于空间的直和分解有唯一的分解式，从而又可推出 $P_i x=y_i=\alpha_i x_i$。将 A 作用到 $x=\sum\limits_{i=1}^{n}\alpha_i x_i$ 的两端，又有

$$Ax = \sum_{i=1}^{n} \alpha_i A x_i = \sum_{i=1}^{n} \lambda_i \alpha_i x_i$$

$$= \sum_{i=1}^{n} \lambda_i P_i x = \left(\sum_{i=1}^{n} \lambda_i P_i \right) x$$

再一次用到$x \in \mathbf{C}^n$的任意性，便得到关系式

$$A = \sum_{i=1}^{n} \lambda_i P_i$$

这说明可以用投影算子 P_i（专指的）及特征值 λ_i（有时叫做A的谱值）来表示A，此即A的谱分解式。当前，我们是在A可对角化的前提下推出的。这里的投影算子 P_i（也是一个复方阵）有下述性质：

1. 幂等性：$P_i^2 = P_i$；
2. 分离性：$P_i P_j = 0$，$i \neq j$；
3. 可加性：$\sum_{i=1}^{n} P_i = I$。

把一个可对角化的算子A分解为一系列投影算子 P_i 的加权和，无论从代数上，还是从几何上进行研究，都有它的方便之处。特别是用$A = \sum_{i=1}^{n} \lambda_i P_i$ 来推导 A^{-1}（若存在的话）及 $f(A)$（A的多项式或解析函数）的谱分解式，尤为方便。为了深入地讨论，我们从幂等算子（或矩阵）谈起。

定义2.3.1 矩阵$P \in \mathbf{C}^{n \times n}$是幂等的，是指它满足

$$P^2 = P$$

引理2.3.2 若$P \in \mathbf{C}^{n \times n}$是幂等的，则有

1. P^H 与$(I-P)$也是幂等的；
2. P的特征值非零即1，且是可对角化的；
3. $\operatorname{rank}(P) = \operatorname{tr}(P)$；
4. $P(I-P) = (I-P)P = 0$；
5. $Px = x \Longleftrightarrow x \in \mathscr{R}(P)$；
6. $\mathscr{N}(P) = \mathscr{R}(I-P)$。

证明 这里只证明2，因为其余几条或据幂等阵的定义，或据2，直接可以推出。设P的秩为r，我们来证明$P\sim \mathrm{diag}(I_r, 0)$。为此，先证明$\mathscr{R}(P)$是$P$关于特征值1的特征子空间，$\mathscr{N}(P)$是$P$关于零特征值的特征子空间；然后再证明$\mathbf{C}^n=\mathscr{R}(P)\dot{+}\mathscr{N}(P)$；最后构造满秩阵$T$，使得

$$T^{-1}PT=\mathrm{diag}(I_r, 0)$$

任取$x\in\mathscr{R}(P)$，则存在$y\in\mathbf{C}^n$使得$x=Py$。于是有$Px=P^2y=Py=x$。这表明$\mathscr{R}(P)$中的每个非零向量都是P关于特征值1的特征向量。反之，若$x=Px$，则显然有$x\in\mathscr{R}(P)$，这表明P关于特征值1的特征向量都属于$\mathscr{R}(P)$。同理可证$\mathscr{N}(P)$是关于零特征值的特征子空间。任取$x\in\mathbf{C}^n$，都有

$$x=Px+(I-P)x，且Px\in\mathscr{R}(P)，(I-P)x\in\mathscr{N}(P)$$

这表明$\mathbf{C}^n=\mathscr{R}(P)+\mathscr{N}(P)$；在第一章证明过，把$P: \mathbf{C}^n\to\mathbf{C}^n$视为一个线性变换，有

$$\dim\mathbf{C}^n=\dim\mathscr{R}(P)+\dim\mathscr{N}(P)$$

故而证明了$\mathbf{C}^n=\mathscr{R}(P)\dot{+}\mathscr{N}(P)$。最后选取$\mathscr{R}(P)$的一个基底$X_1=(x_1x_2\cdots x_r)$及$\mathscr{N}(P)$的一个基底$X_2=(x_{r+1}x_{r+2}\cdots x_n)$，则$(X_1X_2)=(x_1x_2\cdots x_n)$便构成$\mathbf{C}^n$的一个基底，令$T=(X_1\ X_2)\in\mathbf{C}_n^{n\times n}$，则有

$$PT=P(X_1\ X_2)=(PX_1\ PX_2)=(X_1\ 0)$$
$$=(X_1\ X_2)\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$$

此即$T^{-1}PT=\mathrm{diag}(I_r, 0)$。

下面的定理给出了幂等矩阵与投影算子之间的一一对应。设$\mathbf{C}^n=L\dot{+}M$，我们把沿M向L的投影算子记为$P_{L,M}$，即$P_{L,M}$的核为M，值域为L，且$P_{L,M}$限制在L上相当于恒等算子。

定理2.3.3 对任一幂等矩阵$P\in\mathbf{C}^{n\times n}$，$\mathscr{R}(P)$与$\mathscr{N}(P)$是互

补的子空间，即$\mathbf{C}^n=\mathscr{R}(P)\dot{+}\mathscr{N}(P)$，并且$P$是沿$\mathscr{N}(P)$向$\mathscr{R}(P)$的一个投影算子，即$P=P_{R(P),N(P)}$。反之，若$L$与$M$是互补的子空间，则一定存在幂等矩阵$P_{L,M}$，使得$\mathscr{R}(P_{L,M})=L$，$\mathscr{N}(P_{L,M})=M$。

证明 定理的前半部分的证明已包含在定理2.3.2的证明之中。现只证明后半部分。若L，M是互补的，不妨设$\{x_1, x_2, \cdots, x_r\}$与$\{y_1, y_2, \cdots, y_t\}$分别为$L$，$M$的基底，则$\{x_1, x_2, \cdots x_r, y_1, y_2, \cdots, y_t\}$构成$\mathbf{C}^n$的一个基底。现构造$P_{L,M}$如下：据$P_{L,M}$的定义有

$$P_{L,M}x_i=x_i,\quad i=1, 2, \cdots, r$$

$$P_{L,M}y_j=0,\quad j=1, 2, \cdots, t$$

记$X(x_1x_2\cdots x_r)\in\mathbf{C}_r^{n\times r}$，$Y=(y_1y_2\cdots y_t)\in\mathbf{C}_t^{n\times t}$

则$(XY)\in\mathbf{C}_n^{n\times n}$。从而上述方程组等价于

$$P_{L,M}(XY)=(P_{L,M}XP_{L,M}Y)=(X\quad 0)$$

从而可解得

$$P_{L,M}=(X0)(XY)^{-1}$$

构造过程本身很容易验证：对任意的$z\in\mathbf{C}^n$，都有$P_{L,M}^2z=P_{L,M}z$。从而说明$P_{L,M}$是一个幂等阵。

推论2.3.4 设$\mathbf{C}^n=L\dot{+}M$，且$x\in\mathbf{C}^n$的分解式为

$$x=y+z,\quad y\in L,\quad z\in M$$

则$y=P_{L,M}x$，$z=(I-P_{L,M})x$。

推论2.3.5 设$\mathbf{C}^n=L\dot{+}M$，$A\in\mathbf{C}^{n\times n}$。则有

1. $P_{L,M}A=A\Longleftrightarrow\mathscr{R}(A)\subset L$；
2. $AP_{L,M}=A\Longleftrightarrow\mathscr{N}(A)\supset M$；
3. $P_{L,M}=I-P_{M,L}$。

定理2.3.6（可对角化矩阵的谱分解） 设$A\in\mathbf{C}^{n\times n}$有$k$个互异的特征值$\lambda_1, \lambda_2, \cdots, \lambda_k$。则$A$可对角化的充分必要条件是：存在$k$个投影算子$P_1, P_2, \cdots, P_k$，满足

1. $P_i P_j = 0$，$i \neq j$；

2. $\sum\limits_{i=1}^{k} P_i = I$；

3. $A = \sum\limits_{i=1}^{k} \lambda_i P_i$。

证明 充分性。设$r_i = \mathrm{rank}(P_i)$，且令$X_i \in \mathbf{C}_{r_i}^{n \times r_i}$是$\mathscr{R}(P_i)$的某一基构成的列满秩阵，并记

$$X = (X_1 X_2 \cdots X_k)$$

根据引理2.3.2中的3，可知X的列数为

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} \mathrm{rank}(P_i) = \sum_{i=1}^{k} \mathrm{tr}(P_i)$$

$$= \mathrm{tr}\left(\sum_{i=1}^{n} P_i\right) = \mathrm{tr}(I) = n$$

这表明$X \in \mathbf{C}^{n \times n}$。由于$X_i$是$\mathscr{R}(P_i)$的基底构成的矩阵，所以存在$Y_i \in \mathbf{C}^{r_i \times n}$使得$P_i = X_i Y_i$。现记

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{pmatrix}$$

则有

$$XY = (X_1 X_2 \cdots X_k) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{pmatrix} = \sum_{i=1}^{k} X_i Y_i$$

$$= \sum_{i=1}^{k} P_i = I$$

这表明X是一个可逆矩阵。又据引理2.3.2中5有$P_i X_i = X_i$，从而有$P_j X_i = P_j P_i X_i = 0$，$i \neq j$；

$$AX = \left(\sum_{i=1}^{k} \lambda_i P_i\right) X = \sum_{i=1}^{k} \lambda_i (P_i X)$$

$$=(\lambda_1 X_1 \ \lambda_2 X_2 \cdots \lambda_k X_k)$$

$$=(X_1 X_2 \cdots X_k)\begin{pmatrix} \lambda_1 I_{r_1} & & & \\ & \lambda_2 I_{r_2} & & \\ & & \ddots & \\ & & & \lambda_k I_{r_k} \end{pmatrix}$$

$$\triangleq X\Lambda$$

其中 $\Lambda=\mathrm{diag}(\lambda_1 I_{r_1},\ \lambda_2 I_{r_2},\ \cdots,\ \lambda_k I_{r_k})$，此即

$$X^{-1}AX=\Lambda$$

必要性。如果 A 可对角化，即存在满秩阵 X，使得 $AX=X\Lambda$，且 $\Lambda=\mathrm{diag}(\lambda_1 I_{r_1},\ \lambda_2 I_{r_2},\ \cdots,\ \lambda_k I_{r_k})$（适当的调整 X 的列向量的排列顺序，使对应同一特征值 λ_i 的特征向量靠在一起）。再将 X 进行列分块为 $X=(X_1 X_2 \cdots X_k)$，$X_i\in \mathbf{C}_{r_i}^{n\times r_i}$，再对 X^{-1} 进行相应的行分块：

$$X^{-1}=\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{pmatrix},\quad Y_i\in \mathbf{C}_{r_i}^{r_i\times n}$$

并令 $P_i=X_iY_i\in \mathbf{C}^{n\times n}$，则有

$$I=XX^{-1}=\sum_{i=1}^{k}X_iY_i=\sum_{i=1}^{k}P_i$$

$$I=X^{-1}X=\begin{pmatrix} Y_1X_1 & Y_1X_2 & \cdots & Y_1X_k \\ Y_2X_1 & Y_2X_2 & \cdots & Y_2X_k \\ & & \vdots & \\ Y_kX_1 & Y_kX_2 & \cdots & Y_kX_k \end{pmatrix}$$

$$Y_iX_i=I_{r_i},\ Y_iX_j=0,\ i\neq j,$$

从而可推出

$$P_i^2=X_iY_iX_iY_i=X_iY_i=P_i$$

$$P_i\cdot P_j=X_iY_iX_jY_j=X_i0Y_j=0,\ i\neq j$$

$$A = X\Lambda X^{-1} = \sum_{i=1}^{k}\lambda_i X_i Y_i = \sum_{i=1}^{k}\lambda_i P_i$$

推论2.3.7 设可对角化的矩阵 $A\in \mathbf{C}^{n\times n}$ 有 k个互异的特征值λ_1，λ_2，…，λ_k。则

1. A的谱分解式$A=\sum_{i=1}^{k}\lambda_i P_i$ 中的幂等矩阵 P_i 是由A唯一确定的；

2. $P_i = \varphi_i(A)/\varphi_i(\lambda_i)$，$i=1, 2, \cdots, k$

$$\varphi_i(\lambda) = \prod_{\substack{j=1\\ j\neq i}}^{k}(\lambda-\lambda_j)$$

3. 若$f(\lambda)$是任一多项式，则有

$$f(A) = \sum_{i=1}^{k} f(\lambda_i)P_i;$$

4. 矩阵B与A可交换$\Longleftrightarrow B$与P_i可交换。

证明 1. 设$\{Q_1,Q_2,\cdots,Q_k\}$也满足$Q_i^2=Q_i, Q_iQ_j=0$，$i\neq j$，$\sum_{i=1}^{k}Q_i=I$，$A=\sum_{i=1}^{k}\lambda_i Q_i$；则易验证有

$$\begin{matrix} P_iA = AP_i = \lambda_i P_i \\ Q_iA = AQ_i = \lambda_i Q_i \end{matrix} \quad i=1, 2, \cdots, k$$

从而可以得到

$$P_i(AQ_j) = \lambda_j P_i Q_j, \ (P_iA)Q_j = \lambda_i P_i Q_j$$

当$i\neq j$时，$\lambda_i\neq\lambda_j$，故有$P_iQ_j=0$。从而有

$$P_i = P_i\sum_{j=1}^{k}Q_j = P_iQ_i = \Big(\sum_{j=1}^{k}P_j\Big)Q_i = Q_i$$

这便证明了P_i 是由A唯一确定的。

2. 设$G_i=\varphi_i(A)/\varphi_i(\lambda_i)$，$\{P_1, P_2, \cdots, P_k\}$为$A$的谱分解式中的投影算子（简称为谱算子）。则对任意的i，$j=1$，2，…，k，都有

$$G_iP_j = \frac{1}{\varphi_i(\lambda_i)}\prod_{\substack{l=1\\ l\neq i}}^{k}(A-\lambda_l I)P_j$$

$$= \frac{1}{\varphi_i(\lambda_i)} \prod_{\substack{l=1\\ l\neq i}}^{k} (\lambda_j - \lambda_l) P_j$$

$$= [\varphi_i(\lambda_j)/\varphi_i(\lambda_i)] P_j = \delta_{ij} P_j$$

故而有

$$G_i = G_i \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} G_i P_j = \sum_{j=1}^{k} \delta_{ij} P_j = P_i$$

3． 直接从定理2.3.6推出。

4． 由表达式 $A = \sum_{i=1}^{k} \lambda_i P_i$ 及 $P_i = \varphi_i(A)/\varphi_i(\lambda_i)$ 很容易证明，因为B与A可换，也与A的任一多项式阵可换。

如果空间的直和分解是正交直和分解，即

$$\mathbf{C}^n = L \dot{+} M \text{且} M = L^{\perp}$$

则这时投影算子 $P_{L,M}$ 便是向子空间 L 上的正交投影（或垂直投影）。正交投影是最为重要的一种投影，下一章要经常用到它，所以给出如下的一个定理。

定理2.3.8 $P \in \mathbf{C}^{n\times n}$为一个正交投影的充分必要条件是$P$满足：$P^2 = P$及$P^H = P$。

证明 必要性。设 $\mathbf{C}^n = L \dot{+} M$且$M = L^{\perp}$，$P$是沿$M$向$L$上的正交投影。据定理2.3.3知，$P^2 = P$且$\mathscr{R}(P) = L$，$\mathscr{N}(P) = M$。现据$M = L^{\perp}$来证$P^H = P$。为此，任取$x_1$，$x_2 \in \mathbf{C}^n$，则有

$$x_i = y_i + z_i, \quad y_i \in L, \quad z_i \in M,$$

$$(y_i | z_j) = 0, \quad i, \quad j = 1, 2,$$

从而便有

$$(P x_1 | x_2) = (y_1 | y_2 + z_2) = (y_1 | y_2)$$

$$(P^H x_1 | x_2) = (x_1 | P x_2) = (y_1 + z_1 | y_2) = (y_1 | y_2)$$

故有$(P x_1 | x_2) = (P^H x_1 | x_2)$，即$((P - P^H) x_1 | x_2) = 0$。由于该式对任意的 $x_2 \in \mathbf{C}^n$ 都成立，特别取 $x_2 = (P - P^H) x_1$时也成立。这便得出

$$((P - P^H) x_1 | (P - P^H) x_1)$$

$$=\|(P-P^H)x_1\|^2=0$$

根据范数的正定性得到$(P-P^H)x_1=0$。再据$x_1\in\mathbf{C}^n$的任意性，证明了$P-P^H$是定义在$\mathbf{C}^n$上的零算子，从而证明了$P=P^H$。

充分性。由于$P^2=P$，据定理2.3.3知，P是沿着$M=\mathscr{N}(P)$向$L=\mathscr{R}(P)$上的投影。现在只须根据$P=P^H$证明$M=L^\perp$，便证明了P是一个正交投影。为此，任取$x\in\mathscr{N}(P)$及$y\in\mathscr{R}(P)$，则有$Px=0$，$y=Pu$，$u\in\mathbf{C}^n$。从而有

$$(y|x)=(Pu|x)=(u|P^Hx)=(u|Px)=0$$

这表明$\mathscr{N}(P)\perp\mathscr{R}(P)$，即$M=L^\perp$。

有了上述定理，我们可以给出正规矩阵的谱分解式，它的谱算子都是正交投影算子。

定理2.3.9（正规矩阵的谱分解） 设$A\in\mathbf{C}^{n\times n}$有$k$个互异的特征值$\lambda_1$，$\lambda_2$，$\cdots\lambda_k$。则$A$是正规矩阵的充分必要条件为：存在$k$个正交投影算子$P_1$，$P_2$，$\cdots$，$P_k$满足

1. $P_iP_j=0$，$i\neq j$；

2. $\sum_{i=1}^{k}P_i=I$；

3. $A=\sum_{i=1}^{k}\lambda_iP_i$。

证明 充分性。据1及3，再利用$P_i^2=P_i$，$P_i^H=P_i$，便可推得：

$$AA^H=\left(\sum_{i=1}^{k}\lambda_iP_i\right)\left(\sum_{j=1}^{k}\bar{\lambda}_jP_j\right)=\sum_{i=1}^{k}|\lambda_i|^2P_i=A^HA$$

这表明A是正规矩阵。

必要性。因为每个正规矩阵A都酉相似于对角阵，即存在酉阵$U\in\mathbf{C}^{n\times n}$使得

$$A=U\Lambda U^H,\quad \Lambda=\mathrm{diag}(\lambda_1I_{r_1},\ \lambda_2I_{r_2},\ \cdots,\ \lambda_kI_{r_k})$$

现记$U=(X_1X_2\cdots X_k)$，$X_i\in\mathbf{C}_{r_i}^{n\times r_i}$，$P_i=X_iX_i^H$，

则有

$$A=(X_1X_2\cdots X_k)\begin{pmatrix}\lambda_1 I_{r_1} & & & \\ & \lambda_2 I_{r_2} & & \\ & & \ddots & \\ & & & \lambda_k I_{r_k}\end{pmatrix}\begin{pmatrix}X_1^H\\ X_2^H\\ \vdots\\ X_k^H\end{pmatrix}$$

$$=\sum_{i=1}^{k}\lambda_i X_i X_i^H=\sum_{i=1}^{k}\lambda_i P_i$$

及$P_i^H=P_i$。此即定理中的3，及P_i是Hermite阵。又据

$$I=UU^H=(X_1X_2\cdots X_k)(X_1X_2\cdots X_k)^H=\sum_{i=1}^{k}P_i$$

再据

$$I=U^HU=(X_1X_2\cdots X_k)^H(X_1X_2\cdots X_k)$$
$$=[(X_i^HX_j)]$$

可得 $X_i^HX_j=0$，$i\neq j$，$X_i^HX_i=I_{r_i}$，由此易证

$$P_i^2=P_i,\quad P_iP_j=0,\quad i=1,\ 2,\ \cdots,\ k。$$

这便证明了P_1，P_2，…，P_k就是满足定理中1，2，3的正交投影算子组。

第四节　矩阵的Hermite标准形及秩分解

矩阵的Hermite标准形及满秩分解，在解线性方程组及求矩阵的广义逆时，都十分有用，这里作一简单的介绍，为后面讨论作些准备。

定义2.4.1　矩阵$A\in \mathbf{C}_r^{m\times n}$的Hermite标准形$H$应满足：

1．H的前r行中，每行至少含有一个非零元，而后$m-r$行只含零元。

2．H的第i（$i=1,\ 2,\ \cdots,\ r$）行的第一个非零元取1，且出现在第n_i列中，并且$n_1<n_2<\cdots<n_r\leqslant n$。

3．H的第n_i($i=1,\ 2,\ \cdots,\ r$)列中的唯一非零元取1，并且落在第（i，n_i）的位置。

例如矩阵

$$H=\begin{bmatrix} 0 & 1 & \frac{1}{2} & 0 & 1-2i & -\frac{1}{2}i \\ 0 & 0 & 0 & 1 & 2 & 1+i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

便是某一矩阵 $A\in \mathbf{C}_2^{3\times 6}$ 的一个 Hermite 标准形，定义2.4.1 中的 $n_1=2$，$n_2=4$。

很显然，任意的矩阵 $A\in \mathbf{C}_r^{m\times n}$ 总存在非奇异的矩阵P，使得PA为一个 Hermite 标准形。事实上，只要对A作一系列的初等行变换就可达到。所以P是一系列初等变换阵的乘积。如果令$K=[A, I_m]$，作初等行变换后化成如下形式：

$$PK=[PA, P]=[H, P]$$

就把P记录下来。其实是一个消元的过程。

如果进一步把A的Hermite 标准形H的第n_1，n_2，…，n_r列依次调换到前 r 列的位置，则有更简单的形式：

$$\begin{bmatrix} I_r & S \\ 0 & 0 \end{bmatrix}$$

这相当于在$H=PA$的右侧乘上置换阵Q，使得

$$PAQ=\begin{bmatrix} I_r & S \\ 0 & 0 \end{bmatrix}$$

当我们对上面的矩阵再进行初等列变换，又得到A的相抵标准形

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

这表明，对任一矩阵$A\in \mathbf{C}_r^{m\times n}$，都存在满秩阵$P\in \mathbf{C}^{m\times m}$，$Q\in \mathbf{C}^{n\times n}$，使得

$$PAQ=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

且通过对矩阵

$$\begin{bmatrix} A & I_m \\ I_n & 0 \end{bmatrix}$$

的前m行仅施行初等行变换,对前n列仅施行初等列变换,就可化成

$$\begin{pmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} & P \\ Q & 0 \end{pmatrix}$$

从而记录到P及Q。

有了上述的说明，就可给出矩阵的一个满秩分解。

定理2.4.2 对任一$A\in \mathbf{C}_r^{m\times n}$，都存在$B\in \mathbf{C}_r^{m\times r}$，$C\in \mathbf{C}_r^{r\times n}$，使得

$$A=BC$$

称之为A的一个满秩分解。

证明 据前述，存在$P\in \mathbf{C}_m^{m\times m}$，$Q\in \mathbf{C}_n^{n\times n}$，使得

$$A=P^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}Q^{-1}$$

$$=P^{-1}\begin{bmatrix} I_r \\ 0 \end{bmatrix}[I_r \quad 0]Q^{-1}$$

令 $$B=P^{-1}\begin{bmatrix} I_r \\ 0 \end{bmatrix},\quad C=[I_r \quad 0]Q^{-1}$$

则显然$B\in \mathbf{C}_r^{m\times r}$,它是取的$P^{-1}$的前$r$列；$C\in \mathbf{C}_r^{r\times n}$,它是取的$Q^{-1}$的前$r$行。从而证明了$A=BC$。

其实也可以给予定理2.4.2一个抽象的证明：因为rank$(A)=r$，故A有r个线性无关的列向量,记之为$B=(b_1b_2\cdots b_r)\in \mathbf{C}_r^{m\times r}$，而$A$的所有列向量皆可用之线性表出，则存在$C\in \mathbf{C}^{r\times n}$，使得$A=BC$。如果能证明rank$(C)=r$，则表明它就是$A$的一个秩分解。这个证明是很简单的，因为有

$$r\geqslant \text{rank}(C)\geqslant \text{rank}(BC)=\text{rank}(A)=r$$

在这里顺便给出$A\in \mathbf{C}_r^{m\times n}$的Hermite标准形$H$在线性方程

$Ax=b$中的一个应用。设$PA=H$，H是A的Hermite标准形，P是一系列初等变换阵的乘积。则方程 $Ax=b$ 显然等价于 $Hx=Pb$。这时便可得出结论：

1. 若$(Pb)_t\neq 0$，$r+1\leqslant t\leqslant m$，则$Ax=b$是不相容的。

2. 若$(Pb)_t=0$，$r+1\leqslant t\leqslant m$，则 $Ax=b$ 是相容的，且$Ax=0$的解空间$\mathscr{N}(A)=\mathscr{N}(H)$的维数，即$n-r$。特别当

$$H=\begin{pmatrix} & & n_1 & & n_2 & & n_3 & & n_r & \\ 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * & 0 & *\cdots * & 0 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * & 0 & *\cdots * \\ 0\cdots 0 & 0 & 0\cdots 0 & 0 & 0\cdots 0 & 1 & *\cdots * & 0 & *\cdots * \\ \vdots & \vdots & & \vdots & & 0 & & \vdots & \\ & & & & & \vdots & & & \\ & & & & & 0 & 0\cdots 0 & 1 & *\cdots * \\ & & & & & \vdots & & \vdots & \vdots \\ 0\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots 0 \end{pmatrix}$$

时，从第t个方程可解出 x_{n_t}，$1\leqslant t\leqslant r$，它是用其余$n-r$个未知数x_j（$j\neq n_1, n_2, \cdots, n_r$）表示。如果记$J=\{n_1, n_2, \cdots, n_r\}$，则便有

$$x_{n_t}=\sum_{j\notin J}c_{tj}x_j,\quad 1\leqslant t\leqslant r$$

从而$Ax=0$的一般解为

$$\begin{aligned} x&=(x_1, x_2, x_3, \cdots, x_n) \\ &=(x_1, \cdots, x_{n_1}, \cdots, x_{n_2}, \cdots, x_{n_r}, \cdots, x_n) \\ &=(x_1, \cdots, \sum_{j\notin J}c_{1j}x_j, \cdots, \sum_{j\notin J}c_{2j}x_j, \cdots, \\ &\quad \sum_{j\notin J}c_{rj}x_j, \cdots, x_n) \end{aligned}$$

在上述x中，分别令$x_j=0$，$j\in J$，便可得到$n-r$个向量u_1，u_2，$\cdots$，$u_{n-r}\in\mathbf{C}^n$，而x可用之线性表出，且$\{u_1, u_2, \cdots, u_{n-r}\}$是线性无关的。所以它们是方程$Ax=0$的一个基础解系，

即$\mathscr{N}(A)$的一个基底。

3．为了得到$Ax=b$的一个特解，只要在方程$Hx=Pb$中先对x_j（$j\in J$）取任意特定的值，然后关于x_{n_t}（$1\leqslant t\leqslant r$）解出。

定理2.4.2表明，$A\in\mathbf{C}_r^{m\times n}$可以分解为列满秩阵$B$（有的书上叫做高矩阵）和行满秩阵$C$的乘积。而这两种矩阵有一些特殊的性质。为了给出一个重要的性质，先证明一个常用的结论。

引理2.4.3 对任意的$A\in\mathbf{C}_r^{m\times n}$，都有

1．A^HA与AA^H都是半正定的；

2．$\mathrm{rank}(A^HA)=\mathrm{rank}(AA^H)=\mathrm{rank}(A)=r$

证明 1．$\forall x\in\mathbf{C}^n$，及$y\in\mathbf{C}^m$，都有

$$x^HA^HAx=(Ax|Ax)=\|Ax\|^2\geqslant 0$$
$$y^HAA^Hy=(A^Hy|A^Hy)=\|A^Hy\|^2\geqslant 0$$

这便证明了1。

对于2，只证$\mathrm{rank}(A^HA)=\mathrm{rank}(A)$。为此，只要证明方程$Ax=0$与$A^HAx=0$同解就行了，因为它们的解空间的维数分别为$n-\mathrm{rank}(A)$及$n-\mathrm{rank}(A^HA)$。方程$Ax=0$的解显然都也是方程$A^HAx=0$的解。现证方程$A^HAx=0$的解$x_0$也是$Ax=0$的解。因为有

$$0=x_0^HA^HAx_0=(Ax_0)^H(Ax_0)=\|Ax_0\|^2$$

故据范数的正定性有$Ax_0=0$，从而说明x_0也是$Ax=0$的解。引理证完。

根据引理有如下结论：

推论2.4.4 对于$A\in\mathbf{C}^{m\times n}$，$A^HA=0\Longleftrightarrow A=0$。

推论2.4.5 若$A\in\mathbf{C}_r^{m\times n}$的秩分解为$A=BC$，则显然矩阵$(B^HB)^{-1}$与$(CC^H)^{-1}$是存在的，且满足

$$(B^HB)^{-1}B^HB=I_r,\quad CC^H(CC^H)^{-1}=I_r$$

这时，我们把$(B^HB)^{-1}B^H$与$C^H(CC^H)^{-1}$分别叫做B的左逆和C的右逆。

值得注意的是，B与C的左右逆不是唯一的，但上面形式的左右逆特别有用。

如果B是列满秩，且方程 $Bx=c$ 是相容的，则解x是唯一的；如果B是行满秩的，则方程$Bx=c$对任意的 c 都是相容的。这也是列满秩与行满秩阵的重要性质。

第五节　矩阵的奇異值分解和极分解

前面的定理2.1.4说明，任一$A\in \mathbf{C}_r^{m\times n}$，都有分解式

$$A=U\begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix}V^H$$

其中U，V为酉阵，而 $R\in \mathbf{C}_r^{r\times r}$ 是一个正线上三角阵。本节将对上述形式进一步简化，使R为一个对角阵，且主对角元取A的奇异值，从而得到A的奇异值分解。所谓矩阵A的奇异值就是方阵A^HA(或AA^H))的特征值的算术根。它与Hermite 矩阵的特征值有某些相似之处；特别当A是正定的Hermite矩阵时，奇异值就是特征值。根据A的奇异值分解可得到更广意义下的谱分解。它在线性动态系统的辨识、实验数据的处理、以及在正交变换保持不变的范数意义下，所处理的最佳逼近中，都有直接的应用。作为一个例子，这里将用它处理数值分析中的一个带有普遍性的问题，即退化的最小二乘问题。顺便还用它推出矩阵的极分解。

定义2.5.1　对于 $A\in \mathbf{C}^{m\times n}$，非负定矩阵 A^HA 的 n 个特征值 $\lambda_i(\geqslant 0)$的算术根$\sigma_i=\sqrt{\lambda_i}$，叫做$A$的奇异值。

定理2.5.2　设$A\in \mathbf{C}_r^{m\times n}$则存在酉阵$U$及$V$，使得

$$U^HAU=\begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$$

其中$S=\mathrm{diag}\ (\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_r)$ 且$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$。

证明　由于A^HA非负定的Hermite矩阵，其秩为r，故可令其特征值为σ_1^2，σ_2^2，…，σ_n^2，且满足$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$，

$\sigma_{r+1}=\sigma_{r+2}=\cdots=\sigma_n=0$，设$A^HA$关于$\sigma_1^2$，$\sigma_2^2,\cdots,\sigma_n^2$的一组标准正交的特征向量为$v_1$，$v_2$，$\cdots$，$v_n$，并记$V_1=(v_1v_2\cdots v_r)$，$V_2=(v_{r+1}$，$\cdots$，$v_n)$则$V=(V_1V_2)$为酉阵且满足

$$V^HA^HAV=\begin{bmatrix}V_1^H\\V_2^H\end{bmatrix}A^HA(V_1V_2)$$

$$=\begin{bmatrix}V_1^HA^HAV_1 & V_1A^HAV_2\\V_2A^HAV_1 & V_2^HA^HAV_2\end{bmatrix}$$

$$=\begin{bmatrix}S_2 & 0\\0 & 0\end{bmatrix}$$

从而得到

$$V_1^HA^HAV_1=S^2,\ V_2^HA^HAV_2=0$$

上面的第一个等式变形后有

$$S^{-1}V_1^HA^HAV_1S^{-1}=I_r$$

而对第二个等式应用推论2.4.4，便有

$$AV_2=0$$

令$U_1=AV_1S^{-1}\in\mathbf{C}^{m\times r}$，则有$U^HU_1=I_r$。这表明$U_1$是部分酉阵，即它的$r$个列向量是标准正交的。取$U_2\in\mathbf{C}_{m-r}^{m\times(m-r)}$使$U=(U_1U_2)\in\mathbf{C}^{m\times m}$为一酉阵，其实，只要取$\mathscr{R}(U_1)^\perp$的任一标准正交基底便可构成$U_2$。这样便有$U_2^HU_1=0$，从而据上述的推导便有

$$U^HAV=\begin{bmatrix}U_1^H\\U_2^H\end{bmatrix}A(V_1V_2)=\begin{bmatrix}U_1^HAV_1 & U_1^HAV_2\\U_2^HAV_1 & U_2^HAV_2\end{bmatrix}$$

$$=\begin{bmatrix}S^{-1}V_1^HA^HAV_1 & 0\\U_2^HU_1S & 0\end{bmatrix}=\begin{bmatrix}S & 0\\0 & 0\end{bmatrix}$$

这便给出了$A=U\begin{bmatrix}S & 0\\0 & 0\end{bmatrix}V^H$。

上面定理也可以在A的正交分解的基础上给出另外的证明，留作思考。

从定理2.5.2的证明过程可以看出，V的列向量是A^HA的标准正交特征向量，叫做A的右奇异向量；V的前r列是对应着A^HA的r个非零特征值的特征向量。从$U_1=AV_1S^{-1}$的选取可推出

$$
\begin{aligned}
AA^HU_1 &= AA^HAV_1S^{-1}=AV_1S^2S^{-1} \\
&= AV_1S=U_1S^2
\end{aligned}
$$

这表明U_1的r个列向量恰是AA^H关于特征值σ_1^2，σ_2^2，…，σ_r^2的标准正交特征向量。从而说明了A^HA与AA^H有相同的非零特征值，特征向量之间有如下关系

$$u_i=\sigma_i^{-1}Av_i,\quad 1\leqslant i\leqslant r$$

所以又把U的列向量，即AA^H的特征向量叫做A的左奇异向量。从$AV_1=U_1S$，$AV_2=0$还可看出如下的结果：

1. V_2的列向量为$\mathscr{N}(A)$的一个标准正交基底；
2. V_1的列向量为$\mathscr{N}(A)^\perp$的一个标准正交基底；
3. U_1的列向量为$\mathscr{R}(A)$的一个标准正交基底；
4. U_2的列向量为$\mathscr{R}(A)^\perp$的一个标准正交基底；

从奇异值的定义及前面的有关约定，还可看出

5. Hermite阵($A^H=A$)的奇异值σ_i为其特征值λ_i的绝对值，即$\sigma_i=|\lambda_i|$，$i=1,2,\cdots,n$；

6. 若令$\|A\|_2=\max\{\|Ax\|_2/\|x\|_2 \mid x\in\mathbf{C}^n, x\neq0\}$，则有$\|A\|_2=\sigma_1$；

7. 若令$\|A\|_F=\text{tr}(A^HA)$，则$\|A\|_F^2=\sum\limits_{i=1}^n\sigma_i^2$；

根据6及不等式$\|AB\|_2\leqslant\|A\|_2\|B\|_2$，$\|A+B\|_2\leqslant\|A\|_2+\|B\|_2$又可推出：

8. $\sigma_1(AB)\leqslant\sigma_1(A)\sigma_1(B)$，$\sigma_1(A+B)\leqslant\sigma_1(A)+\sigma_1(B)$，其中$\sigma_1(A)$代表$A$的最大奇异值。

定理2.5.3 设$A\in\mathbf{C}^{m\times n}$的奇异值为$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant0$，

则有

$$\sigma_i = \min_{\dim(S)=n-i+1} \max_{\substack{x\in S \\ x\neq 0}} \frac{\| Ax \|_2}{\| x \|_2}$$

其中S为$\mathbf{C}^n$的子空间，$\| x \|_2 = (x^H x)^{1/2}$。

证明 设$\{u_1, u_2, \cdots, u_n\}$及$\{v_1, v_2, \cdots, v_n\}$分别为$A$对应于奇异值$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$的左右奇异向量，则它们分别构成酉空间$\mathbf{C}^m$及$\mathbf{C}^n$的标准正交基底，并且有$A = \sum_{i=1}^{n} \sigma_i u_i v_i^H$。很显然，对任一$n-i+1$维子空间$S$，一定存在非零向量$x$满足

$$x \perp \text{span}(v_{i+1}, v_{i+2}, \cdots, v_n)$$

从而有

$$x = \sum_{j=1}^{i} \gamma_j v_j, \quad \| x \|_2 = \left(\sum_{j=1}^{i} |\gamma_j|^2 \right)^{1/2}$$

$$Ax = \left(\sum_{i=1}^{n} \sigma_i u_i v_i^H \right) \left(\sum_{j=1}^{i} \gamma_j v_j \right) = \sum_{j=1}^{i} \sigma_j \gamma_j u_j$$

$$\| Ax \|_2 = \left(\sum_{j=1}^{i} |\sigma_j \gamma_j|^2 \right)^{1/2} \geqslant \sigma_i \left(\sum_{j=1}^{i} |\gamma_j|^2 \right)^{1/2}$$

$$(\| Ax \|_2 / \| x \|_2) \geqslant \sigma_i$$

上面不等式的两端关于所有的$n-i+1$维子空间S及所有非零的向量x取极小值，仍有

$$\min_{\substack{x\in S \\ x\neq 0}} (\| Ax \|_2 / \| x \|_2) \geqslant \sigma_i$$

为了证明上式右端能取到最小值σ_i，只须构造一个特殊的$n-i+1$维子空间$S' = \text{span}(v_i, v_{i+1}, \cdots, v_n)$。对于任一非零的$x \in S'$，显然有

$$x = \sum_{j=i}^{n} \gamma_j v_j, \quad \| x \|_2 = \left(\sum_{j=i}^{n} |\gamma_j|^2 \right)^{1/2}$$

$$\| Ax \|_2 = \left(\sum_{j=i}^{n} \sigma_j^2 |\gamma_j|^2 \right)^{1/2} \leqslant \sigma_i \left(\sum_{j=i}^{n} |\gamma_j|^2 \right)^{1/2}$$

$$(\| Ax \|_2 / \| x \|_2) \leqslant \sigma_i$$

当然有
$$\max_{\substack{x\in S'\\x\neq 0}}(\|Ax\|_2/\|x\|_2)\leqslant\sigma_i$$

$$\min_{\substack{x\in S'\\x\neq 0}}(\|Ax\|_2/\|x\|_2)\geqslant\sigma_i$$

从而证明了在S'上可以达到极小值σ_i。亦即
$$\min_{\dim S=n-i+1}\max_{\substack{x\in S\\x\neq 0}}(\|Ax\|_2/\|x\|_2)=\sigma_i$$

定理2.5.4 设 $A, B\in \mathbf{C}^{m\times n}$ 分别具有奇异值$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n$及$\tau_1\geqslant\tau_2\geqslant\cdots\geqslant\tau_n$。则$\sigma_i$与$\tau_i$之间有如下的关系
$$|\sigma_i-\tau_i|\leqslant\|A-B\|_2,\ i=1,\ 2,\ \cdots,\ n$$

证明 设 $A=B+E$，S' 为使$\max\limits_{\substack{x\in S\\x\neq 0}}(\|Ax\|_2/\|x\|_2)$取极小的 $n-i+1$ 维子空间。则有
$$\sigma_i=\max_{\substack{x\in S'\\x\neq 0}}\frac{\|Bx+Ex\|_2}{\|x\|_2}$$
$$\leqslant\max_{\substack{x\in S'\\x\neq 0}}\frac{\|Bx\|_2}{\|x\|_2}+\max_{\substack{x\in S'\\x\neq 0}}\frac{\|Ex\|_2}{\|x\|_2}$$
$$\leqslant\tau_i+\|E\|_2=\tau_i+\|A-B\|_2$$
类似地可得到
$$\tau_i\leqslant\sigma_i+\|A-B\|_2$$
两个不等式联立起来，便有
$$|\sigma_i-\tau_i|\leqslant\|A-B\|_2$$

这个定理表明，当矩阵A有微小摄动而成为$A_1=A+E$时，则它的奇异值的改变不会大于摄动矩阵的2-范数。

定理2.5.5 设 $A\in\mathbf{C}_l^{m\times n}$，$m\geqslant n$，$l\geqslant r$，$A$的奇异值为$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$；记$S=\mathrm{diag}(\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_n)$，$\sum=\mathrm{diag}(\sigma_1,\ \sigma_2,\ \cdots,\ \sigma_r)$，$A$ 的奇异值分解为

$$A=U\begin{bmatrix}S\\0\end{bmatrix}V^H$$

现记

$$A_0=U\begin{bmatrix}\Sigma & 0\\0 & 0\end{bmatrix}V^H\text{，}\operatorname{rank}(A_0)=\operatorname{rank}(\Sigma)$$

则有

$$\|A-A_0\|_F=\min_{\operatorname{rank}(B)=r}\|A-B\|_F$$

这个定理表明在Frobenius范数意义下，所有秩为r的$m\times n$阶矩阵中，A_0是A的一个最佳逼近。

证明 我们首先证明，如果A的最佳逼近B_0存在的话，则一定有

$$\|A-B_0\|_F^2=\sum_{i=r+1}^{n}\sigma_i^2\text{；}$$

然后再证明我们构造的A_0确实满足

$$\|A-A_0\|_F^2=\sum_{i=r+1}^{n}\sigma_i^2$$

为此，设$B_0\in\mathbf{C}_r^{m\times n}$，使$\|A-B\|_F$达到极小，即

$$\|A-B_0\|_F=\min\{\|A-B\|_F\mid B\in\mathbf{C}_r^{m\times n}\}$$

并设B_0的奇异值分解为

$$B_0=Q\begin{bmatrix}C_{11} & 0\\0 & 0\end{bmatrix}P^H$$

其中Q，P为酉阵，$C_{11}=\operatorname{diag}(\gamma_1,\gamma_2,\cdots,\gamma_r)$的主对角元为$B_0$的$r$个非零奇异值。作矩阵

$$C=Q^HB_0P=\begin{bmatrix}C_{11} & 0\\0 & 0\end{bmatrix},\quad D=Q^HAP=\begin{bmatrix}D_{11} & D_{12}\\D_{21} & D_{22}\end{bmatrix}$$

则可以断言：$D_{11}=C_{11}$，$D_{12}=0$，$D_{21}=0$；否则，以$D_{12}\neq0$为例，作矩阵

$$M=\begin{bmatrix}C_{11} & D_{12}\\0 & 0\end{bmatrix}\in\mathbf{C}_r^{m\times n}$$

则有

$$\| D-M \|_F < \| D-C \|_F$$

再令 $N=QMP^H$，便有

$$\| A-N \|_F = \| D-M \|_F < \| D-C \|_F = \| A-B_0 \|_F$$

这显然与所设 $\| A-B_0 \|_F$的极小性矛盾。类似地可证 $D_{11}=C_{11}$，$D_{21}=0$。据此可得

$$D=\begin{bmatrix} C_{11} & 0 \\ 0 & D_{22} \end{bmatrix}$$

$$\| A-B_0 \|_F = \| D-C \|_F = \| D_{22} \|_F$$

注意到A与D有相同的奇异值，而C_{11}的奇异值与D_{22}的奇异值的并集为A的奇异值。又C_{11}的奇异值恰是A的前r个奇异值，故D_{22}的奇异值只能为A的后$n-r$个奇异值，从而得到

$$\| D_{22} \|_F^2 = \sigma_{r+1}^2 + \cdots + \sigma_n^2 = \sum_{i=r+1}^{n} \sigma_i^2$$

至于$\| A-A_0 \|_F^2 = \sum_{i=r+1}^{n} \sigma_i^2$是显然的。所以$A_0$是$A$的一个秩为$r$的最佳矩阵逼近。

定理2.5.6 设$A\in \mathbf{C}_r^{m\times n}$的奇异值分解为

$$A=U\begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}V^H$$

$b\in \mathbf{C}^m$是一个特定的向量。$f(x)=\| b-Ax \|_2$是定义在整个空间$\mathbf{C}^n$上的非负实值函数。$\| \cdot \|_2$是取酉空间$\mathbf{C}^m$中的范数。则向量

$$x_0=V\begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}U^H b$$

是使函数$f(x)$达到极小的向量；如果$f(x)$的极小向量不唯一，则x_0是其中具有最小范数的极小向量。也就是说x_0是最小二乘问题

$$\min_{x\in C^n} \| b - Ax \|_2$$

的最小范数解。

证明 令$z=V^H x$，$c=U^H b$，且作适当的分块：

$$z=\begin{bmatrix} z_1 \\ z_2 \end{bmatrix},\quad c=\begin{bmatrix} c_1 \\ c_2 \end{bmatrix},\quad z_1,\ c_1\in \mathbf{C}^r$$

则有

$$\begin{aligned}
\| b-Ax \|_2^2 &= \| U^H(b-Ax) \|_2^2 \\
&= \| U^H b - U^H AVV^H x \|_2^2 \\
&= \left\| c - \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} z \right\|_2^2 = \left\| \begin{pmatrix} c_1 - Sz_1 \\ c_2 \end{pmatrix} \right\|_2^2 \\
&= \| c_1 - Sz_1 \|_2^2 + \| c_2 \|_2^2 \geqslant \| c_2 \|_2^2
\end{aligned}$$

这表明对一切$x\in \mathbf{C}^n$，函数$f(x)$有下界$\| c_2 \|_2$，并且在取$z_1=S^{-1}c_1$，z_2任意时，$f(x)$可以达到这个下界。从而也说明

$$x=Vz=V\begin{bmatrix} S^{-1}c_1 \\ z_2 \end{bmatrix},\quad \forall z_2\in \mathbf{C}^{n-r}$$

是$f(x)$的极小点。并且很容易看出

$$x=V\begin{bmatrix} S^{-1}c_1 \\ 0 \end{bmatrix}=V\begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}=V\begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}U^H b$$

是$f(x)$的极小点中具有最小范数者。

定理2.5.7（极分解） 设$A\in \mathbf{C}^{n\times n}$，则$A$可分解为

$$A=GU=UH$$

其中G，H为半正定的Hermite阵，U为酉阵。

证明 把A的奇异值分解作适当的变形：

$$\begin{aligned}
A &= U_1\begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}V_1^H \\
&= U_1\begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}U_1^H U_1 V_1^H
\end{aligned}$$

$$= U_1 V^H V_1 \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} V_1^H$$

并记

$$G = U_1 \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} U_1^H, \quad H = V_1 \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} V_1^H, \quad U = U_1 V_1^H$$

则便有

$$A = GU = UH$$

显然G，H为半正定的Hermite阵，U为酉阵。

*第六节 Jordan标准形的存在性

方阵$A \in \mathbf{C}^{n \times n}$的Jordan标准形是相似意义下的最简形式，它刻划了A的很多特征。若记

$$P^{-1}AP = \mathrm{diag}(J_1, J_2, \cdots, J_m)$$

并把P的列向量对应的分块为

$$P = (P_1 P_2 \cdots P_m)$$

即使P_i的列数与J_i的阶数相等，则有

$$AP_i = P_i J_i, \qquad i = 1, 2, \cdots, m$$

$$\mathbf{C}^n = \mathscr{R}(P_1) \dot{+} \mathscr{R}(P_2) \dot{+} \cdots \dot{+} \mathscr{R}(P_m)$$

这表明A在$\mathbf{C}^n$上的作用可以简化为它在$\mathscr{R}(P_i)$上的作用来分析，并且每个P_i的第一列都是A的一个特征向量。在第四章中将会看到，Jordan标准形在矩阵分析中是一个很重要的工具。但是要证明Jordan标准形的存在性及具体构造出来，无论是用λ-矩阵的理论，还是用循环不变子空间的理论，都格外的复杂，有时甚至构造不出来。在第二节证明舒尔引理后，我们曾指出可用Jordan标准形的存在性及满秩阵的UR分解也能证明舒尔引理。现在要问，借助舒尔引理能否也可证明Jordan标准形的存在性呢？下面就是给出的一个回答。

定理2.6.1 给定$A\in\mathbf{C}^{n\times n}$，则一定存在非奇异阵$X\in\mathbf{C}^{n\times n}$，使得

$$X^{-1}AX=\mathrm{diag}(J_1,\ J_2,\ \cdots,\ J_s)$$

其中

$$J_i=\lambda_i I_{k_i}+E_i,\quad E_i=\begin{pmatrix}0 & I_{k_i-1}\\ 0 & 0\end{pmatrix},\ i=1,\ 2,\ \cdots,\ s$$

证明 本定理的证明分以下三步：

1． 根据舒尔引理，对每个$B\in\mathbf{C}^{n\times n}$，都存在酉阵$U$，使得$B=URU^H$，其中$R$为上三角阵，并且还可以要求把$R$的主对角元按相同者衔接排列起来。

2． 构造非奇异阵X，使得

$$X^{-1}RX=\mathrm{diag}(R_1,\ R_2,\ \cdots,\ R_m)$$

其中

$R_j=\lambda_j I+U_j$，U_j为严格上三角阵，且诸λ_j互异。

3． 证明每个R_j都有一个Jordan 标准形。

我们采用引理的形式给出上面的过程的实现，最后作一个综合。

引理2.6.2 设$R_1\in\mathbf{C}^{k_1\times k_1}$，$R_2\in\mathbf{C}^{k_2\times k_2}$是上三角阵，$S\in\mathbf{C}^{k_1\times k_2}$，则矩阵方程

$$R_1B-BR_2=S$$

有唯一解$B\in\mathbf{C}^{k_1\times k_2}$的充要条件是

$$\sigma(R_1)\cap\sigma(R_2)=\phi$$

其中$\sigma(R)$表示R的特征值构成的集合。

证明 若我们把矩阵

$$A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}\in\mathbf{C}^{m\times n}$$

的拉直定义为

$$\overrightarrow{A}=(a_{11}a_{12}\cdots a_{1n}a_{21}a_{22}\cdots a_{2n}\cdots a_{m1}a_{m2}\cdots a_{mn})^T$$

则根据矩阵乘法和加法的运算，矩阵方程 $R_1B-BR_2=S$ 等价于线性方程组

$$[R_1\otimes I_{k_2}-I_{k_1}\otimes R_2^T]\overrightarrow{B}=\overrightarrow{S}$$

其中$A\otimes B$表示矩阵A与B的直积，或Kronecker积：

$$A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ & \vdots & & \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

设$\sigma(R_1)=\{\lambda_1,\lambda_2,\cdots,\lambda_{k_1}\}$，$\sigma(R_2)=\{\mu_1,\mu_2,\cdots,\mu_{k_2}\}$，根据直积的定义及$R_1$，$R_2$分别为上三角阵的特点，很容易看出有

$$R_1\otimes I_{k_2}+I_{k_1}\otimes(-R_2^T)$$

$$=\begin{pmatrix} \lambda_1 I_{k_2}-R_2^T & & & * \\ & \lambda_2 I_{k_2}-R_2^T & & \\ & & \ddots & \\ & & & \lambda_{k_1} I_{k_2}-R_2^T \end{pmatrix}$$

从而有

$$\det[R_1\otimes I_{k_2}+I_{k_1}\otimes(-R_2^T)]=\prod_{i=1}^{k_1}\det(\lambda_i I_{k_2}-R_2^T)$$

$$=\prod_{i=1}^{k_1}\prod_{j=1}^{k_2}(\lambda_i-\mu_j)$$

这便证明了定理的结论。

引理2.6.3 设 $R\in\mathbf{C}^{n\times n}$ 是一个上三角阵，则存在非奇异阵$X\in\mathbf{C}^{n\times n}$，使得

$$X^{-1}RX=\mathrm{diag}(R_1,R_2,\cdots,R_n)$$

其中$R_i=\lambda_i I+U_i$，U_i为严格上三角阵，λ_i互异。

证明 对n进行归纳证明。当$n=1$时，命题显然成立。设对于阶数小于n的上三角阵R_1及K命题为真，现对$R\in\mathbf{C}^{n\times n}$为上三角阵的情况证明。据舒尔的分解过程，不妨设上三角阵

$$R=\begin{bmatrix} R_1 & S \\ 0 & K \end{bmatrix}$$

其中R_1与K是没有公共特征值的上三角阵，且$R_1=\lambda_1 I+U_1$，U_1为严格上三角阵。这时要想保证有适当阶数的矩阵B，使得等式

$$\begin{bmatrix} I & B \\ 0 & I \end{bmatrix}\begin{bmatrix} R_1 & S \\ 0 & K \end{bmatrix}\begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}=\begin{bmatrix} R_1 & 0 \\ 0 & K \end{bmatrix}$$

成立，其充分必要条件是

$$S=R_1B-BK$$

而据假设$\sigma(R_1)\cap\sigma(K)=\phi$及引理2.6.2知，上述方程有唯一的$B$存在。从而据归纳假设，存在适当阶数的矩阵$C$，使得

$$C^{-1}KC=\mathrm{diag}(R_2,\ R_3,\ \cdots,\ R_m)$$

从而取

$$X=\begin{bmatrix} I & -B \\ 0 & I \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & C \end{bmatrix}=\begin{bmatrix} I & -BC \\ 0 & C \end{bmatrix}$$

便有

$$\begin{aligned} X^{-1}RX &= \begin{bmatrix} I & B \\ 0 & C^{-1} \end{bmatrix}\begin{bmatrix} R_1 & S \\ 0 & K \end{bmatrix}\begin{bmatrix} I & -BC \\ 0 & C \end{bmatrix} \\ &= \begin{bmatrix} R_1 & SC-R_1BC+BKC \\ 0 & C^{-1}KC \end{bmatrix} \\ &= \begin{pmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_m \end{pmatrix} \end{aligned}$$

引理2.6.4 设$E\in\mathbf{C}^{k\times k}$且$E=\begin{pmatrix}0 & I_{k-1}\\0 & 0\end{pmatrix}$，则

$$E^TE=\begin{bmatrix}0 & 0\\0 & I_{k-1}\end{bmatrix},\qquad E^k=0,\ Ee_{i+1}=e_i,$$

$$(I-E^TE)r=\sigma e_1,\qquad \sigma=e_1^Tr,$$

其中$e_i\in\mathbf{C}^k$且第i个分量为1，其余分量皆为零。

证明 直接验证便知结论正确。

引理2.6.5 设$U\in\mathbf{C}^{n\times n}$且为严格上三角阵，则存在非奇异阵$X\in\mathbf{C}^{n\times n}$使得

$$X^{-1}UX=J$$

其中

$$J=\mathrm{diag}(E_1,\ E_2\cdots,\ E_m),$$

$$E_j=\begin{pmatrix}0 & I_{k_j}\\0 & 0\end{pmatrix}=\begin{bmatrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{bmatrix}$$

（E_{j+1}的阶数）$\leqslant$（E_j的阶数）$=k_j+1$，$j=1,\ 2,\ \cdots,\ m-1$。

如果把J写出，则有

$$J=\begin{pmatrix}E_1 & & & \\ & E_2 & & \\ & & \ddots & \\ & & & E_m\end{pmatrix}=\begin{pmatrix}\boxed{\begin{matrix}0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1\\ & & & 0\end{matrix}} & & & \\ & \boxed{\begin{matrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{matrix}} & & \\ & & \ddots & \\ & & & \boxed{\begin{matrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{matrix}}\end{pmatrix}$$

证明 对n进行归纳证明。当$n=1$时，$U=0$，命题显然成

立。假定对阶数小于 n 的严格上三角阵命题为真。现对严格上三角阵 $U\in\mathbf{C}^{n\times n}$ 加以证明。先将 U 分块为

$$U=\begin{bmatrix}0 & u^T\\ 0 & U_1\end{bmatrix},\quad u\in\mathbf{C}^{n-1}$$

根据归纳假设，存在非奇异阵 X_1，使得

$$X_1^{-1}U_1X_1=\begin{bmatrix}E_1 & 0\\ 0 & J_1\end{bmatrix}=\mathrm{diag}(E_1,\ E_2,\ \cdots,\ E_m)$$

$J_1=\mathrm{diag}(E_2,\ \cdots,\ E_m)$，（$E_1$ 的阶数）$\geqslant$（E_j 的阶数），$j\geqslant 2$。从而有

$$\begin{aligned}&\begin{bmatrix}1 & 0\\ 0 & X_1^{-1}\end{bmatrix}U\begin{bmatrix}1 & 0\\ 0 & X_1\end{bmatrix}\\ &=\begin{bmatrix}1 & 0\\ 0 & X_1^{-1}\end{bmatrix}\begin{bmatrix}0 & u^T\\ 0 & U_1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & X_1\end{bmatrix}\\ &=\begin{bmatrix}0 & u^TX_1\\ 0 & X_1^{-1}U_1X_1\end{bmatrix}\end{aligned}$$

再将 u^TX_1 列分块成 $u^TX_1=(u_1^T\ \ u_2^T)$，使得与 $X_1^{-1}U_1X_1$ 中的列分块保持一致。这样便有

$$\begin{bmatrix}0 & u^TX_1\\ 0 & X_1^{-1}U_1X_1\end{bmatrix}=\begin{pmatrix}0 & u_1^T & u_2^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}$$

借助引理 4，可以推得

$$\begin{aligned}&\begin{pmatrix}1 & -u_1^TE_1^T & 0\\ 0 & I & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix}0 & u_1^T & u_2^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}\begin{pmatrix}1 & u_1^TE_1^T & 0\\ 0 & I & 0\\ 0 & 0 & I\end{pmatrix}\\ &=\begin{pmatrix}0 & u_1^T-u_1^TE_1^TE_1 & u_2^T\\ 0 & E_1 & 0\\ 0 & 0 & 0\end{pmatrix}=\begin{pmatrix}0 & \sigma e_1^T & s^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}\end{aligned}$$

下面对 σ 分两种情况讨论：

如果$\sigma\neq0$，则有

$$\begin{pmatrix}\sigma^{-1} & 0 & 0\\ 0 & I & 0\\ 0 & 0 & \sigma^{-1}I\end{pmatrix}\begin{pmatrix}0 & \sigma e_1^T & s^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}\begin{pmatrix}\sigma & 0 & 0\\ 0 & I & 0\\ 0 & 0 & \sigma I\end{pmatrix}$$

$$=\begin{pmatrix}0 & e_1^T & \sigma^{-1}s^T\\ 0 & E_1 & 0\\ 0 & 0 & \sigma^{-1}J_1\end{pmatrix}\begin{pmatrix}\sigma & 0 & 0\\ 0 & I & 0\\ 0 & 0 & \sigma I\end{pmatrix}=\begin{pmatrix}0 & e_1^T & s^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}$$

$$=\begin{bmatrix}E & e_1s^T\\ 0 & J_1\end{bmatrix}$$

即
$$U\sim\begin{bmatrix}E & e_1s^T\\ 0 & J_1\end{bmatrix}\quad E=\begin{bmatrix}0 & e_1^T\\ 0 & E_1\end{bmatrix}$$

$$(E\text{的阶数})=(E_1\text{的阶数})+1$$

因为E的阶数严格大于J_1中的每个对角块的阶数，故$J_1^{k-1}=0$。现定义

$$s_i^T=s^TJ_1^{i-1},\quad i=1,\ 2,\ \cdots,\ k,$$

注意到$s=u_2$，故有

$$\begin{bmatrix}I & e_{i+1}s_i^T\\ 0 & I\end{bmatrix}\begin{bmatrix}E & e_is_i^T\\ 0 & J_1\end{bmatrix}\begin{bmatrix}I & -e_{i+1}s_i^T\\ 0 & I\end{bmatrix}$$

$$=\begin{bmatrix}E & -Ee_{i+1}s_i^T+e_is_i+e_{i+1}s_i^TJ_1\\ 0 & J_1\end{bmatrix}$$

$$=\begin{bmatrix}E & e_{i+1}s_{i+1}^T\\ 0 & J_1\end{bmatrix},\qquad 1\leqslant i\leqslant k-1$$

因为$J_1^{k-1}=0$，$s_k=(J_1^{k-1})^Ts=0$，$s_1=s$，所以在上式中取$i=1$，2，…，$k-1$，便有

$$U\sim\begin{bmatrix}E & e_1s_1^T\\ 0 & J_1\end{bmatrix}\sim\begin{bmatrix}E & e_2s_2^T\\ 0 & J_1\end{bmatrix}\sim\cdots\sim\begin{bmatrix}E & e_ks_k^T\\ 0 & J_1\end{bmatrix}=\begin{bmatrix}E & 0\\ 0 & J_1\end{bmatrix}$$

这样便证明了

$$U\sim\mathrm{diag}(E_1,\ E_2,\ \cdots,\ E_m)$$

如果$\sigma=0$，则有

$$U\sim\begin{pmatrix}0 & 0 & s^T\\ 0 & E_1 & 0\\ 0 & 0 & J_1\end{pmatrix}$$

由于 E_1 位于主对角块上，故可作行与列相对应的置换，使之变为

$$U\sim\begin{pmatrix}E_1 & 0 & 0\\ 0 & 0 & s^T\\ 0 & 0 & J_1\end{pmatrix}$$

由归纳假设，对严格上三角阵$\begin{bmatrix}0 & s^T\\ 0 & J_1\end{bmatrix}$，存在非奇异阵 X_2 使得

$$X_2^{-1}\begin{bmatrix}0 & s^T\\ 0 & J_1\end{bmatrix}X_2=J_2=\mathrm{diag}(E_2',\ \cdots,\ E_m')$$

于是再对诸对角块根据其阶次大小作适当置换，便得到命题所需要的形式。

综合起来，给定$A\in\mathbf{C}^{n\times n}$，其 Jordan 标准形的推导过程是：

第一步：据舒尔引理，存在酉阵X_1，使得

$$X_1^{-1}AX_1=R$$

其中上三角阵R的主对角元按相同者衔接排列起来，舒尔引理的证明过程表明可以这样做。

第二步：据引理2.6.3，存在非奇异阵X_2，使得

$$X_2^{-1}RX_2=\mathrm{diag}(R_1,\ R_2,\ \cdots,\ R_m)$$

其中

$$R_j=\lambda_jI+U_j,\ \lambda_j\neq\lambda_i,\ U_j\text{为严格上三角阵。}$$

第三步：据引理2.6.5，对每个严格上三角阵U_j，存在非奇异阵Y_j，使得

$$Y_j^{-1}U_jY_j=\mathrm{diag}(E_{j1},\ E_{j2},\ \cdots,\ E_{jm_j})$$

$$Y_j^{-1}R_jY_j=\lambda_jI+Y_j^{-1}U_jY_j$$
$$=\lambda_jI+\mathrm{diag}(E_{j1},\ E_{j2},\ \cdots,\ E_{jm_j})$$

$$\begin{pmatrix} Y_1^{-1} & & \\ & \ddots & \\ & & Y_m^{-1}\end{pmatrix}\begin{pmatrix} R_1 & & \\ & \ddots & \\ & & R_m\end{pmatrix}\begin{pmatrix} Y_1 & & \\ & \ddots & \\ & & Y_m\end{pmatrix}$$
$$=\begin{pmatrix}\lambda_1I+\mathrm{diag}(E_{11},E_{12},\cdots,E_{1m_1}) & & \\ & \ddots & \\ & & \lambda_mI+\mathrm{diag}(E_{m1},\cdots,E_{mm_m})\end{pmatrix}$$

第四步：令$X_3=\mathrm{diag}(Y_1,\ Y_2,\ \cdots,\ Y_m)$，则有

$$X_3^{-1}X_2^{-1}X_1^{-1}AX_1X_2X_3$$
$$=X_3^{-1}X_2^{-1}RX_2X_3=X_3^{-1}\mathrm{diag}(R_1,\ R_2,\ \cdots,\ R_m)X_3$$
$$=\mathrm{diag}(J_1,\ J_2\cdots,\ J_m)$$

其中$J_i=\mathrm{diag}(J_{i1},\ J_{i2},\ \cdots,\ J_{im_i})$，$i=1,\ 2,\ \cdots,\ m$

$$J_{il}=\lambda_iI+E_{il},\quad l=1,\ 2,\ \cdots,\ m_i$$

注意，最后证明的结果反映出对A的同一特征值λ_i可以有几个小的 Jordan 块。但这在定理 2.6.1 中没有强调，而是允许λ_i（$i=1,\ 2,\ \cdots,\ s$）可以有相同者，所以下脚s表示 Jorban 块的个数。

第七节　例题与习题

1．试证：方阵$A\in\mathbf{R}^{n\times n}$为对称正定的充分必要条件是$A^{-1}$也是对称正定的。

2．若$A\in\mathbf{R}^{n\times n}$是对称阵。试证存在正实数$\mu$，使得$\mu I+A$为正定，$-\mu I+A$为负定。

3．若半定对称阵A，$B\in\mathbf{R}^{n\times n}$可交换，试证$AB$也是半定对称阵。

4．试证：方阵A是正规的充分必要条件为A可以分解成：

$A=SQ=QS$，其中Q为正交阵或酉阵，S为半正定对称阵或半正定 Hermite 阵。

5. 试证：实对称方阵A的特征值全部落在区间$[a, b]$上的充分必要条件是$A-aI$与$bI-A$都为半正定对称阵。

6. 若实对称方阵A与B的特征值分别落在区间$[a, b]$与$[c, d]$内，证明对称阵$A+B$的特征值一定落在区间$[a+c, b+d]$内。

7. （极分解式）任一n阶非奇异实方阵A，都能唯一地分解为$A=S_1Q=QS_2$，其中Q为正交阵，S_1与S_2为正定对称阵。

提示：A^TA为正定对称阵，故它的算术平方根S亦为正定对称阵，从而有

$$I=S^{-1}S^2S^{-1}=S^{-1}(A^TA)S^{-1}=(AS^{-1})^T(AS^{-1})$$

这表明AS^{-1}是一个正交阵，记之为$Q=AS^{-1}$,从而有$A=QS$，唯一性据S的唯一性可证。

8. 根据习题7的结论证明：对每个n阶非奇异实方阵A，都存在正交阵Q_3及Q_4，使得

$$Q_3AQ_4=\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

其中$0<\lambda_n\leqslant\lambda_{n-1}\leqslant\cdots\leqslant\lambda_1$，且$\{\lambda_i^2\}_{i=1}^n$为$AA^T$的所有特征值。

提示：设A的极分解式为$A=SQ_1$，其中$S^2=AA^T$。由于S是正定对称阵，故它正交相似于对角阵，即存在正交阵Q_3，使得

$$Q_3SQ_3^T=\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

亦即 $$S=Q_3^T\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)Q_3$$

从而有

$$A=SQ_1=Q_3^T\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)Q_3Q_1$$

令$Q_4=(Q_3Q_1)^T$,便得到$A=Q_3^T\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)Q_4^T$。这个结论是说：每个非奇异的实方阵$A$都正交相抵于一个对角阵，且

对角阵的主对角元为矩阵 AA^T 的特征值的算术根。这是奇异值分解式的特殊情况。不过它有很重要的几何意义：

n 维仿射空间 $\mathbf{R}^n$ 中的每一个仿射变换：

$$y=Ax+b,\ \det(A)\neq 0,\ b\in\mathbf{R}^n$$

都可分解为以下几个特殊的仿射变换的乘积：

(1) 平移 $y=x+b$

(2) 正交变换 $y=Qx$，$Q^T=Q^{-1}$

(3) 压缩变换 $y=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}x$

請读者自己验证一下。

9. 若方阵 $A\in\mathbf{R}^{n\times n}$ 是对称的且满足 $A^2=0$，你能用几种方法证明 $A=0$。

10. 试证：对每一个对称实方阵 A，都存在一个对称方阵 S，使得 $S^3=A$。

11. 如果 A 是一个正规矩阵，W 是 A 的一个不变子空间（$AW\subset W$），试证 W 的正交补 $W^\perp$ 也是 A 的不变子空间。

12. 如果 A 关于特征值 λ 的特征向量为 φ，则显然有

$$\lambda=\frac{\varphi^H A\varphi}{\varphi^H\varphi}$$

这表明 λ 等于 A 关于 φ 的 Rayleigh 商。这是用来求特征值的一种方法。用这种表示证明 Hermite 矩阵的特征值都是实的，正定矩阵的特征值都是正的，半正定矩阵的特征值都是非负的。

13. 求矩阵 $A=\begin{pmatrix}1 & 1\\ 4 & 1\end{pmatrix}$ 的谱分解式。

首先求得 A 的特征值为 $\lambda_1=3$，$\lambda_2=-1$，而相应的特征向量为

$$\varphi_1=(1,\ 2)^T,\ \varphi_2=(1,\ -2)^T$$

而矩阵 $A^H=\begin{pmatrix}1 & 4\\1 & 1\end{pmatrix}$的特征值也是3，－1，相应的特征向量为

$$\psi_1=\frac{1}{2}\left(1,\ -\frac{1}{2}\right)^T,\ \psi_2=\frac{1}{2}\left(1,\ -\frac{1}{2}\right)^T$$

则向$E(\lambda_1)$上的投影为

$$P_1=\varphi_1\psi_1^H;\ P_1x=\frac{1}{2}\left(\xi_1+\frac{1}{2}\xi_2\right)\varphi_1$$

而向$E(\lambda_2)$上的投影为

$$P_2=\varphi_2\psi_2^H;\ P_2x=\frac{1}{2}\left(\xi_1-\frac{1}{2}\xi_2\right)\varphi_2$$

其中 $x=(\xi_1,\ \xi_2)^T\in\mathbf{C}^2$，最后便得到

$$A=\lambda_1P_1+\lambda_2P_2=3P_1-P_2$$

如果要问：为什么要通过求A^H的特征向量ψ_i来构造投影算子P_i？道理很简单，因为可对角化矩阵A的谱分解式是根据关系式

$$A=P\Lambda P^{-1}$$

推导出来的，其中$P=(\varphi_1,\ \varphi_2,\ \cdots,\ \varphi_n)$的列向量 φ_i 是A关于特征值λ_i的特征向量，而对角阵

$$\Lambda=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$$

若记$P^{-1}=(\psi_1,\ \psi_2,\ \cdots,\ \psi_n)^H$，则显然有

$$A=P\Lambda P^{-1}=\sum_{i=1}^{n}\lambda_i\varphi_i\psi_i^H=\sum_{i=1}^{n}\lambda_iP_i$$

其中$P_i=\varphi_i\psi_i^H$ 就是平行着 $[E(\lambda_i)]^c=\bigoplus\limits_{\substack{j=1\\j\neq i}}^{n}E(\lambda_j)$ 向 $E(\lambda_i)$上的投影算子。但据$P^{-1}A=\Lambda P^{-1}$有

$$(\psi_1\psi_2\cdots\psi_n)^HA=\Lambda(\psi_1\psi_2\cdots\psi_n)^H$$

两边取共轭转置便有

$$A^H(\psi_1\psi_2\cdots\psi_n)=(\psi_1\psi_2\cdots\psi_n)\Lambda^H$$

从而可得

$$A^H\psi_i=\bar{\lambda}_i\psi_i,\ i=1,2,\cdots,n$$

这表明ψ_i是A^H关于特征值$\bar{\lambda}_i$的特征向量。

14. 证明一个正规矩阵若是三角阵，则一定是对角阵。

15. 证明：正规阵A是幂零阵（$A^2=0$），当且仅当$A=0$。

16. 在欧氏空间 $\mathbf{R}^m$ 中，垂直投影算子是很有用的。如果M是 $\mathbf{R}^m$ 的一个子空间，我们考虑沿$M^\perp$向M上的垂直投影算子$P_{M,M^\perp}$（简记为P_M）是怎样计算的。首先考查一维的$M=\mathrm{span}(a)$，也就是求向直线M上的垂直投影。$\mathbf{R}^m$ 中的内积取标准内积$(x|y)=y^Tx$；向量$x\in\mathbf{R}^m$在a上的投影记为x_a，x与a之间的夹角记为$\theta(\leqslant 90°)$，向量a上的单位向量记为$\omega=a/\|a\|$，则据几何知识，x在a上的投影的长度（模）为

$$\|x\|\cos\theta$$

而投影则为

$$x_a=\omega\|x\|\cos\theta=\frac{a}{\|a\|}\cdot\|x\|\cos\theta$$

由于

$$\cos\theta=\frac{(x|a)}{\|x\|\ \|a\|}$$

故有

$$x_a=(x|\omega)\omega=\omega^Tx\omega=\omega\omega^Tx$$

也就是

$$x_a=P_ax=\omega\omega^Tx$$

从而得到投影算子P_a的表示：

$$P_a=\omega\omega^T=\frac{aa^T}{\|a\|^2}\in\mathbf{R}^{n\times n}$$

很容易验证P_a是幂等的 Hermite 阵，且

$$\mathcal{N}(P_a)=\{a\}^\perp,\quad \mathcal{R}(P_a)=\mathrm{span}(a)$$

从$P_a=\omega\omega^T$来看，P_a只决定于a方向上的单位矢量ω，而与其

它无关。现在考虑M是 $\mathbf{R}^m$ 的 n 维子空间的情况。如果$\{\omega_1, \omega_2, \cdots, \omega_n\}$是$M$的一个标准正交基底，任取 $x\in\mathbf{R}^m$，它在ω_i上的垂直投影为

$$x_{\omega_i}=P_{\omega_i}x=\omega_i\omega_i^T x, \qquad i=1, 2, \cdots, n$$

而 x 在M上的垂直投影为

$$\begin{aligned} P_M x &= P_{\omega_1}x+P_{\omega_2}x+\cdots+P_{\omega_n}x \\ &=(\omega_1\omega_1^T+\omega_2\omega_2^T+\cdots+\omega_n\omega_n^T)x \\ &=(\omega_1\omega_2\cdots\omega_n)(\omega_1\omega_2\cdots\omega_n)^T x \end{aligned}$$

若记$(\omega_1\omega_2\cdots\omega_n)=Q\in\mathbf{R}^{m\times n}$，则$Q$是一个部分酉（正交）阵，即$Q^TQ=I_n$，所以得到

$$\begin{aligned} P_M &= QQ^T \\ &=(\omega_1\omega_2\cdots\omega_n)(\omega_1\omega_2\cdots\omega_n)^T \end{aligned}$$

这表明，给出了M的标准正交基底，求 P_M 也是很容易的。读者自己验证 P_M 是幂等的 Hermite 阵，并且满足$\mathscr{N}(P_M)=M^\perp$，$\mathscr{R}(P_M)=M$。

最后，我们考查给出M的一般基底$\{a_1, a_2, \cdots a_n\}$（$a_i\in\mathbf{R}^m$，$i=1, 2, \cdots, n$）的情形。对该基底施行 Gram-Schmidt标准正交化，得到标准正交基底$\{\omega_1, \omega_2, \cdots, \omega_n\}$。若记

$$A=(a_1a_2\cdots a_n)\in\mathbf{R}_n^{m\times n}$$

$$Q=(\omega_1\omega_2\cdots\omega_n)\in\mathbf{R}_n^{m\times n}, \ (Q^TQ=I_n)$$

则据推论2.1.3后面的说明，可知有如下关系

$$A=QR, \ R\in\mathbf{R}_n^{n\times n}$$

因此有$Q=AR^{-1}$，且 $A^TA=R^TR$。据前面推导可得到

$$\begin{aligned} P_M &= QQ^T=AR^{-1}(R^{-1})^TA^T \\ &=A(R^TR)^{-1}A^T=A(A^TA)^{-1}A^T \end{aligned}$$

不难验证：

$$P_M^2=P_M, \ P_M^H=P_M, \ \mathscr{N}(P_M)=\mathbf{R}(A)^\perp,$$

$$\mathscr{R}(P_M)=M$$

如果知道M是由$\{a_1, a_2, \cdots, a_n\}$线性张成的，而$\{a_1, a_2, \cdots, a_n\}$又是线性相关的，试问$P_M$又如何用矩阵$A=(a_1, a_2, \cdots, a_n)$给出？在学过广义逆矩阵之后，便有了回答。

17． 仿照习题16推导酉空间$\mathbf{C}^n$中的正交投影算子。

18． 现在考虑这样一个问题：在n维欧氏空间$\mathbf{R}^n$中，给定等模的两个向量a，b（$\|a\|=\|b\|$），欲求一个正交变换H：$\mathbf{R}^n \longrightarrow \mathbf{R}^n$，它使得

$$Ha=b$$

我们假定a与b的夹角$\theta \leqslant 90°$。取单位向量

$$\omega=\frac{a-b}{\|a-b\|}$$

则向量a，b，$a-b$构成一个等腰三角形。而a在ω上的投影为

$$a_\omega=\omega\omega^T a$$

从而有

$$a-b=2\omega\omega^T a$$

亦即

$$\begin{aligned} b &= a-2\omega\omega^T a \\ &= (I-2\omega\omega^T)a \end{aligned}$$

这样便得到变换$H=I-2\omega\omega^T$，它显然是把a变为b的一个正交变换，即满足$H^T=H^{-1}$，并且还满足$H=H^{-1}$或$H^2=I$。这说明a经过两次这种变换又还原为自己。如果把a与b夹角的平分面（与$a-b$垂直，当然也与ω垂直）视为一面镜子，则H便是一个镜象反射，又叫做Hauseholder变换，它是数值计算中的一个很重要的变换。读者就$\mathbf{R}^2$中给定$a=(1, 2)^T$，$b=(2,1)^T$的情况，计算出H，并作出几何上的解释。如果令

$$\omega=\frac{a+b}{\|a+b\|}$$

则$H=2\omega\omega^T-I$也满足所提出的要求，试从几何上作出解释。

19. Cayley-Hamilton 定理有一系列的应用，这里考查它在矩阵幂的计算方面的应用例子。给定

$$A=\begin{pmatrix}1 & 2\\ 4 & 3\end{pmatrix}$$

容易算出它的特征多项式为

$$\lambda^2-4\lambda-5=0$$

据C-H定理有

$$A^2-4A-5I=0$$

$$A^2=4A+5I=\begin{pmatrix}9 & 8\\ 16 & 17\end{pmatrix}$$

$$A^3=AA^2=A(4A+5I)=4A^2+5A$$

$$=21A+20I=\begin{pmatrix}41 & 42\\ 84 & 83\end{pmatrix}$$

$$A^{-1}=\frac{1}{5}(A-4I)=\begin{pmatrix}-3/5 & 2/5\\ 4/5 & -1/5\end{pmatrix}$$

$$A^{-2}=-\frac{4}{25}A+\frac{21}{25}I=\begin{pmatrix}17/25 & -8/25\\ -16/25 & 9/25\end{pmatrix}$$

很显然，对任意的正负整数 r，都有

$$A^r=\alpha_1A+\alpha_2I$$

其中α_1，α_2根据 r 的不同有所改变。当然，$r<0$ 时A 应该是非奇异的。一般来说，对于 $A\in\mathbf{C}^{n\times n}$，应该有

$$A^r=\alpha_1A^{n-1}+\alpha_2A^{n-2}+\cdots+\alpha_nI$$

现以$r>0$证之。据多项式带余除法有

$$\lambda^r=f(\lambda)Q(\lambda)+R(\lambda) \qquad (*)$$

其中余式$R(\lambda)$的次数应低于特征多项式$f(\lambda)$的次数，故可令

$$R(\lambda)=\alpha_1\lambda^{n-1}+\cdots+\alpha_n$$

（*）的两端用A代替λ，利用C-H定理$f(A)=0$，便得到

$$A^r=\alpha_1A^{n+1}+\alpha_2A^{n-2}+\cdots+\alpha_nI$$

如果λ_1，λ_2，…，λ_n 为A的特征值，即$f(\lambda_i)=0$，据（*）又

可得到方程组：

$$\lambda_i^r=\alpha_1\lambda_i^{n-1}+\alpha_2\lambda_i^{n-2}+\cdots+\alpha_n,\ i=1,\ 2,\ \cdots,\ n$$

这是含有 n 个未知数 α_1，α_2，…，α_n 的线性方程组。如果 $\{\lambda_i\}_{i=1}^n$ 互不相同，则该方程组的系数行列式是满秩的，因它是λ_1，λ_2，…，λ_n 的范德蒙行列式。这时有唯一的一组系数 α_1，α_2，…α_n。如果特征值 λ_i 是重根（$f(\lambda)=0$的重根），不妨设其重数为m_i（代数重数）。则方程组中有m_i 个方程是一样的。不过这时有

$$f(\lambda)=(\lambda-\lambda_i)^{m_i}p(\lambda),\qquad p(\lambda_i)\neq 0$$

从而可知：

$$f^{(s)}(\lambda_i)=0,\ s=0,\ 1,\ \cdots,\ m_i-1。$$

对（*）两端关于λ求导m_i-1次，用λ_i代入便得到

$$R^{(s)}(\lambda_i)=r(r-1)\cdots(r-s+1)\lambda_i^{r-s},$$

$$s=0,\ 1,\ \cdots,\ m_i-1$$

这样仍有 n 个含有未知数 α_1，α_2，…，α_n 的方程组，并且可以证明该方程有解。兹举例说明：

考虑矩阵

$$A=\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}$$

的 r 次幂A^r的表达式。显然A有特征值$\lambda_1=\lambda_2=1$，又知

$$A^r=\alpha_1A+\alpha_2I$$

对应的多项式为

$$\lambda^r=\alpha_1\lambda+\alpha_2 \tag{1}$$

求导一次又有

$$r\lambda^{r-1}=\alpha_1 \tag{2}$$

令$\lambda=1$代入（1）、（2）便得到

$$\alpha_1+\alpha_2=1,\ \alpha_1=r,\ \alpha_2=1-r$$

从而有

$$A^r=rA+(1-r)I$$

$$A^8=3A-2I=\begin{pmatrix}1 & 6\\ 0 & 1\end{pmatrix}$$

$$A^{26}=26A-25I=\begin{pmatrix}1 & 52\\ 0 & 1\end{pmatrix}$$

$$A^{-1}=-A+2I=\begin{pmatrix}1 & -2\\ 0 & 1\end{pmatrix}$$

20. 已知矩阵

$$A=\begin{pmatrix}1 & 0 & 0 & 0\\ 1 & 1 & 0 & 0\\ 0 & 1 & 1 & 0\\ -1 & -1 & 0 & 1\end{pmatrix}$$

试证：

$$A^r=\begin{pmatrix}1 & 0 & 0 & 0\\ r & 1 & 0 & 0\\ \dfrac{r(r-1)}{2} & r & 1 & 0\\ -\dfrac{r(r+1)}{2} & -r & 0 & 1\end{pmatrix}$$

21. 求矩阵

$$A=\begin{pmatrix}3 & -2 & -4\\ -2 & 6 & -2\\ -4 & -2 & 3\end{pmatrix}$$

的谱分解式，并给出A^n的表达式。

22. 证明：如果一个实对称矩阵A的主对角元都大于零，则A至少有一个正特征值。

23. 若$A\in \mathbf{R}_n^{m\times n}$，证明方阵$C=A^TA$一定存在一个对称正定的平方根，而方阵$B=AA^T$一定存在对称半正定的平方根。

24. 定义在酉空间$\mathbf{C}^n$上的任何一个正交投影$P(P^2=P$且

$P^H=P$）都是半正定的。试证明之。

提示：$x^HPx=x^HP^2x=x^HPPx$
$$=x^HP^HPx=\|Px\|^2\geqslant 0$$

25． 证明：任一非奇异方阵都存在一系列排列阵P，使得方阵PA的顺序主子式都不等于零。

26． 对下列矩阵作满秩的分解

$$A=\begin{pmatrix} 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix},\quad B=\begin{pmatrix} 1 & 2 & 3 & 6 \\ 2 & 4 & 6 & 12 \\ 1 & 2 & 3 & 6 \\ 2 & 4 & 6 & 12 \end{pmatrix}$$

27． 设矩阵

$$A=\begin{pmatrix} -1 & i & 0 \\ -i & 0 & -i \\ 0 & i & -1 \end{pmatrix},\quad B=\begin{pmatrix} 0 & i & 1 \\ -i & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

试问：A与B是正规矩阵吗？若是，通过酉变换把它们化成相似的对角阵。

28． 设A的满秩分解为$A=BC$。试证：
$$Ax=0\Longleftrightarrow Cx=0$$

29． 若$A=(a_{ij})$是n阶正定阵，则有
$$\det(A)\leqslant a_{11}a_{22}\cdots a_{nn}$$
当且仅当A为对角阵有等式成立，这就是有名的哈达玛不等式。

提示：用归纳法只要证明不等式
$$\det(A)\leqslant a_{nn}\det(A_{n-1})$$
当且仅当$a_{1n}=a_{2n}=\cdots=a_{n-1\,n}=0$有等式成立就够了，其中$A_{n-1}$为$A$的$n-1$阶顺序主子式。为此，记
$$A=\begin{bmatrix} A_{n-1} & a \\ a^H & a_{nn} \end{bmatrix}$$

因为

$$\begin{bmatrix} I_{n-1} & 0 \\ -a^H A_{n-1}^{-1} & 1 \end{bmatrix}\begin{bmatrix} A_{n-1} & a \\ a^H & a_{nn} \end{bmatrix}=\begin{bmatrix} A_{n-1} & a \\ 0 & a_{nn}-a^H A_{n-1}^{-1}a \end{bmatrix}$$

从而有

$$\det(A)=(a_{nn}-a^H A_{n-1}^{-1}a)\det(A_{n-1})$$

由于A_{n-1}，A_{n-1}^{-1}都是正定的，故$\det(A_{n-1})>0$，且$a^H A_{n-1}^{-1}a\geqslant 0$，当且仅当$a=(a_{1n}, a_{2n}, \cdots, a_{n-1\,n})^T=0$时$a^H A_{n-1}^{-1}a=0$。
从而有

$$\det(A)\leqslant a_{nn}\det(A_{n-1})$$

当且仅当$a=0$时等式成立。

30. 证明：欧氏空间R^n中的任一超二次曲面

$$f(x_1, x_2, \cdots, x_n)=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j$$

都能通过正交变换

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=\begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ & \vdots & & \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

化成标准形

$$f(x_1, x_2, \cdots, x_n)=\sum_{i=1}^{n}\lambda_i y_i^2$$

提示：先将超二次曲面$f(x_1, x_2, \cdots, x_n)$写成矩阵形式。为此，令

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},\ a_{ij}=a_{ji}=\frac{a_{ij}+a_{ji}}{2}$$

$$x=(x_1, x_2, \cdots, x_n)^T$$

则有

$$f(x_1, x_2, \cdots, x_n)=x^T Ax$$

其中A是 n 阶实对称阵。由于存在正交阵 Q使得

$$A=Q^T\Lambda Q$$

$$\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

故而有

$$f(x)=x^TAx=x^TQ^T\Lambda Qx$$

$$=y^T\Lambda y=\sum_{i=1}^{n}\lambda_i y_i^2$$

其中$y=Qx$。例如

$$f(x_1, x_2, x_3)=2x_1^2+2x_2^2+2x_1x_2+3x_3^2$$

$$=x^TAx$$

其中

$$x=(x_1, x_2, x_3)^T, \quad A=\begin{pmatrix}2 & 1 & 0\\ 1 & 2 & 0\\ 0 & 0 & 3\end{pmatrix}$$

不难求出A之特征值为$\lambda_1=1$，$\lambda_2=\lambda_3=3$，对应的标准正交特征向量为

$$q_1=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{pmatrix}, \quad q_2=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}\\ 0\end{pmatrix}, \quad q_3=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

构造矩阵

$$Q=(q_1q_2q_3)=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0\\ 0 & 0 & 1\end{pmatrix}$$

则有

$$Q^T A Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

作变换

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = Q^T x = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

便得到

$$f(x_1,\ x_2,\ x_3) \overset{y=Q^T x}{=\!=\!=} \sum_{i=1}^{3} \lambda_i y_i^2 = y_1^2 + 3y_2^2 + 3y_3^2$$

它是一个椭球面。

第三章　矩阵的广义逆

大家知道，当$A\in C_n^{n\times n}$时，方程$Ax=b$有唯一的解，并且可表示为$x=A^{-1}b$,结论是如此的简单而美妙。但是在应用领域中却更多地要碰到奇异方阵和长方阵，并且大量的问题，最终要归结到求解算子方程$Ax=b$。人们自然会想到：能否有一个广义的逆矩阵，第一，它具有通常逆矩阵的部分性质，并且在A满秩时，这种广义逆与通常逆吻合；第二，用这种广义逆同样能求解方程$Ax=b$并进行理论分析。这就是提出和研究广义逆矩阵的出发点。

广义的逆的概念似乎是1903年由 Fredholm 首先关于积分算子提出的，他称之为伪逆。而微分算子的广义逆已隐含在Hilbert 关于广义的格林函数的讨论（1904）之中。嗣后的二十多年有很多学者，如Myller（1906）、Westfall（1909）、Bounitzky（1909）、Hurwitz（1912）、Elliott（1928）、Reid（1931）等，都做了不同度程的研究。特别是Hurwitz利用 Fredholm 算子的零空间的有限维性，给予伪逆一个简单的代数构造。对于无穷维算子的广义逆的研究，自然刺激了对有限维线性算子矩阵的广义逆的研究。1920年，E.H.Moore 在美国数学学会的一个会议上，以摘要的形式指出，对每个有限的矩阵，都存在唯一的一个广义逆，他叫做"general reciprocal"，但详细的论文在他死后的1935年才发表。他是用正交投影算子来定义广义逆的，我们把它叫做 Moore 广义逆。1951年，Bjerhammer搞清了Moore逆的最小二乘性质，也就是说，若G是A的Moore逆，则Gb给出了方程$Ax=b$的最小二乘解。1955年，R.Penrose在Bjerhammer 的研究基础上，提出了用方程组$AGA=A$，$GAG=G$，$(AG)^H=AG$，$(GA)^H=GA$来定义A的广义逆G，叫做 Penrose 逆，也证明了G的唯一性。不久，

Bjerhammer发现并证明了 Moore 逆与 Penrose 逆完全是一个数学概念。所以后来把它叫做Moorl-Penrose 逆，并记作A^+。五十年代以后，在这方面出现了大批的研究论文，特别值得指出的数学家有Grevill (1957)、Ben-Israel (1957)、Scroggs和Odell (1966)、Rao和Mitra等。1953年，Bott-Duffin 出于研究网络理论的需要，还提出过所谓方阵的约束广义逆。六十年代以后，有的人从不同的角度，对广义逆的概念又加以推广；有的人则进一步挖掘各种广义逆的性质或构造具体的计算方法；也有人研究广义逆在各个领域的实际应用。目前，广义逆已形成了一套系统而完整的理论，并成为很多研究领域中的有用工具。

本章主要介绍这方面的基本概念和结论，并给出一些应用的例子。

第一节 Moore-Penrose广义逆

定义3.1.1(Moore) 设$A\in \mathbf{C}^{m\times n}$，如果$G\in \mathbf{C}^{n\times m}$满足

$$AG=P_{\mathscr{R}(A),\mathscr{N}(A^H)},\quad GA=P_{\mathscr{R}(A^H),\mathscr{N}(A)}$$

则称G为 A 的 Moore 广义逆；上面等式的右端是两个正交投影算子。

定义3.1.2 (Penrose) 设$A\in \mathbf{C}^{m\times n}$，如果$G\in \mathbf{C}^{n\times m}$满足

1. $AGA=A$
2. $GAG=G$
3. $(AG)^H=AG$
4. $(GA)^H=GA$

则称G为A的Penrose广义逆。

后面将要证明Moore逆与Penrose逆是等价的，并且是唯一的，故称之为$M-P$逆，记为A^+。由于Penrose 的四个方程都各有一定的解释，并且应用起来有方便之处，所以出于不同的目的，常要

考虑满足部分方程的G，叫做弱逆。为引用的方便，给予如下的记法。

定义3.1.3 设$A\in\mathbf{C}^{m\times n}$，用$A\{i,j,\cdots,l\}$表示所有满足Penrose第$i,j,\cdots,l$个方程的矩阵$G\in\mathbf{C}^{n\times m}$的集合；$G\in A\{i,j,\cdots,l\}$又常记作$G=A^{(i,j,\cdots,l)}$，叫做$A$的$\{i,j,\cdots,l\}$-逆。

首先证明$A\{1,2,3,4\}$是单元素集，即A^+的唯一性，这里先承认$A\{1,2,3,4\}\neq\phi$。

设$X,\ Y\in A\{1,2,3,4\}$，则有

$$\begin{aligned}X&=X(AX)^H=XX^HA^H=XX^H(AYA)^H\\&=X(AX)^H(AY)^H=XAY=(XA)^H(YA)^HY\\&=A^HY^HY=(YA)^HY=Y\end{aligned}$$

下面我们从构造$A\{1\}$的元素开始，逐步地构造$A\{1,2\}$，$A\{1,3\}$，$A\{1,4\}$及A^+，并刻画它们的特征。这个过程较长且显得有点繁琐，读者要耐心些。上一章曾谈到：对任一$A\in\mathbf{C}_r^{m\times n}$，存在非奇异矩阵$E\in\mathbf{C}_m^{m\times m}$及置换阵$P\in\mathbf{C}_n^{n\times n}$，使得

$$EAP=\begin{bmatrix}I_r & K\\0 & 0\end{bmatrix}\tag{1}$$

此即A的Hermite标准形。据此有

定理3.1.4 设$A\in\mathbf{C}_r^{m\times n}$，$E\in\mathbf{C}_m^{m\times m}$及$P\in\mathbf{C}_n^{n\times n}$使得式(1)成立。则对任一$L\in\mathbf{C}^{(n-r)\times(m-r)}$，矩阵

$$X=P\begin{bmatrix}I_r & 0\\0 & L\end{bmatrix}E\in\mathbf{C}^{n\times m}\tag{2}$$

是A的一个$\{1\}$-逆，即$X\in A\{1\}$。

证明 因为

$$A=E^{-1}\begin{bmatrix}I_r & K\\0 & 0\end{bmatrix}P^{-1}$$

直接验算便知有$AXA=A$。

当$r=\mathrm{rank}(A)=0$时，任何$n\times m$阶矩阵皆为A的一个$\{1\}$-

逆，因为这时A为零矩阵。从式(2)可以看出 $\mathrm{rank}X=r+\mathrm{rank}(L)$；由于 L 的任意性，故$A\{1\}$包含无穷多个元素，且其秩可任取大于或等于r，而小于或等于$\min(m,n)$的正整数。

对于任一$A^{(1)}\in A\{1\}$，显然有如下性质。

引理3.1.5 设$A\in \mathbf{C}^{m\times n}$，$\lambda\in C$，则有

1. $(A^{(1)})^H\in A^H\{1\}$；
2. 若$A\in \mathbf{C}_n^{n\times n}$，则$A^{(1)}=A^{-1}$，且是唯一的；
3. $\lambda^+A^{(1)}\in(\lambda A)\{1\}$，$\lambda^+=\begin{cases}\lambda^{-1}, & \lambda\neq 0\\ 0, & \lambda=0\end{cases}$；
4. $\mathrm{rank}(A^{(1)})\geqslant\mathrm{rank}(A)$；
5. 若S，T为非奇异矩阵，则$T^{-1}A^{(1)}S^{-1}\in(SAT)\{1\}$；
6. $AA^{(1)}$及$A^{(1)}A$都是幂等阵，且满足
   $\mathrm{rank}(AA^{(1)})=\mathrm{rank}(A^{(1)}A)=\mathrm{rank}(A)$

引理3.1.6 设$A\in \mathbf{C}_r^{m\times n}$，则有

1. $A^{(1)}A=I_n\Longleftrightarrow r=n$；此时$A^{(1)}$叫做$A$的左逆，记为$A_L^{-1}$。
2. $AA^{(1)}=I_m\Longleftrightarrow r=m$；此时$A^{(1)}$叫做$A$的右逆，记为$A_R^{-1}$。

证明 1. ($\Longleftarrow$)若 $A\in \mathbf{C}_n^{m\times n}$，则据引理3.1.5的6知，幂等矩阵$A^{(1)}A$是非奇异的，因此对等式

$$(A^{(1)}A)^2=A^{(1)}A$$

两端乘上$(A^{(1)}A)^{-1}$，便得到$A^{(1)}A=I_n$。

($\Longrightarrow$) 因为$A^{(1)}A=I_n$，则由引理3.1.5的6知

$$\mathrm{rank}(A)=\mathrm{rank}(A^{(1)}A)=\mathrm{rank}(I_n)=n$$

2. 类似对1的证明。

引理3.1.7 若$A^{(1)}\in A\{1\}$，则

$$\mathscr{R}(AA^{(1)})=\mathscr{R}(A),\quad \mathscr{N}(A^{(1)}A)=\mathscr{N}(A),$$
$$\mathscr{R}((A^{(1)}A)^H)=\mathscr{R}(A^H)$$

证明 首先有 $\mathscr{R}(AA^{(1)})\subset\mathscr{R}(A)$，据引理 3.1.5 的 6 有 $\text{rank}(A)=\text{rank}(AA^{(1)})$，故 $\dim\mathscr{R}(A)=\dim\mathscr{R}(AA^{(1)})$，从而 $\mathscr{R}(AA^{(1)})=\mathscr{R}(A)$。

$\mathscr{N}(A)\subset\mathscr{N}(A^{(1)}A)$ 是显然的；据线性方程组的理论，从条件$\text{rank}(A)=\text{rank}(A^{(1)}A)$可推得$\dim\mathscr{N}(A)=\dim\mathscr{N}(A^{(1)}A)$，故有$\mathscr{N}(A)=\mathscr{N}(A^{(1)}A)$。

最后，据引理3.1.5的1，可得

$$\mathscr{R}(A^H)\supset\mathscr{R}(A^H(A^H)^{(1)})$$
$$=\mathscr{R}((A^{(1)}A)^H)\supset\mathscr{R}(A^H(A^{(1)})^H A^H)=\mathscr{R}(A^H)$$

一般情况下，有如下结论：

$$\mathscr{R}(AB)=\mathscr{R}(A)\Longleftrightarrow\text{rank}\,AB=\text{rank}\,A$$
$$\mathscr{N}(AB)=\mathscr{N}(B)\Longleftrightarrow\text{rank}\,AB=\text{rank}\,B$$

請读者作为练习证明之。

根据 $A^{(1)}$ 的存在性，很容易证明$A^{(1,2)}$的存在性， 从而也就证明了$A^{(2)}$的存在性。

引理3.1.8 设Y，$Z\in A\{1\}$， 并令 $X=YAZ$ ，则有 $X\in A\{1,2\}$。

证明 直接可验证X满足Penrose的前两个方程。

由于$G\in A\{1,2\}$与A在Penrose方程1，2中是对称出现的，因此有人又把$\{1,2\}$-逆叫做自反广义逆，即A与G是互为$\{1,2\}$-逆。由于矩阵乘积的秩不得大于每个因子的秩， 故任一$G\in A\{1,2\}$都有与A相同的秩。并且还有

定理3.1.9(Bjerhammer) 给定A及$G\in A\{1\}$，则有

$$G\in A\{1,2\}\Longleftrightarrow\text{rank}(G)=\text{rank}(A)。$$

证明 ($\Longrightarrow$) 是显然的。现证($\Longleftarrow$)。由于$G\in A\{1\}$，故据引理3.1.5的6有$\text{rank}(GA)=\text{rank}(A)$；而据假设 $\text{rank}(G)$，$=\text{rank}(A)$,(故有$\text{rank}(GA)=\text{rank}(G)$。很显然有$\mathscr{R}(GA)\subset\mathscr{R}(G)$，但据 $\text{rank}(GA)=\text{rank}(G)$知 ，$\dim\mathscr{R}(GA)=\dim\mathscr{R}(G)$ ，故有$\mathscr{R}(GA)=\mathscr{R}(G)$。这表明存在矩阵$Y$，使得

$$G = GAY$$

上式两端左乘GA，便有

$$GAG = GAGAY = GAY = G$$

从而证明了$G \in A\{1, 2\}$。

上述定理也给出了与之等价的

推论3.1.10 下面三个结论中的任何两个都蕴含着第三个：

$G \in A\{1\}$，$G \in A\{2\}$，$\text{rank}(G) = \text{rank}(A)$。

从定理3.1.4 中所构造的X 不难发现，若取 $L = 0$ ，便有$\text{rank}(X) = \text{rank}(A) = r$，故而

$$G = P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}E$$

便是A的一个$\{1, 2\}$-逆。

上面我们讨论了 $\{1\}$-逆及$\{1, 2\}$-逆的存在性及构造方法。下面的两个定理将给出$\{1, 2, 3\}$-逆，$\{1, 2, 4\}$-逆及$\{1, 2, 3, 4\}$-逆的存在性及构造方法。

定理3.1.11(**Urguhart**) 对任一$A \in \mathbf{C}^{m \times n}$ ，都有

$$Y = (A^H A)^{(1)} A^H \in A\{1, 2, 3\}$$

$$Z = A^H (AA^H)^{(1)} \in A\{1, 2, 4\}$$

证明 第二章已证过$\text{rank}(A^H A) = \text{rank}(AA^H) = \text{rank}(A^H) = \text{rank}(A)$，再注意到$\mathscr{R}(A^H A) \subset \mathscr{R}(A^H)$，便得到

$$\mathscr{R}(A^H A) = \mathscr{R}(A^H)$$

从而知存在某个矩阵U，使得

$$A^H = A^H AU$$

两边同取共轭转置，便有

$$A = U^H A^H A$$

因而可知

$$AYA = U^H A^H A(A^H A)^{(1)} A^H A = U^H A^H A = A$$

这说明$Y \in A\{1\}$；从而知$\text{rank}(Y) \geqslant \text{rank}(A)$，但从$Y$的定义又知$\text{rank}(Y) \leqslant \text{rank}(A^H) = \text{rank}(A)$，故推出$\text{rank}(Y) = \text{rank}(A)$；从

而据定理3.1.9知

$$Y\in A\{1,\ 2\}$$

最后，据$A^H=A^H AU$及$A=U^H A^H A$可推出

$$\begin{aligned}AY&=U^H A^H A(A^H A)^{(1)}A^H\\&=U^H A^H A(A^H A)^{(1)}A^H AU\\&=U^H A^H AU=(AU)^H(AU)\end{aligned}$$

这表明AY是一个Hermite阵。从而证得$Y\in A\{1,\ 2,\ 3\}$。类似地推理过程，可证明$Z\in A\{1,\ 2,\ 4\}$。

定理3.1.12(Urguhart) 对任一$A\in \mathbf{C}^{m\times n}$，都有

$$A^+=A^{(1,4)}AA^{(1,3)}$$

证明 令$Y=A^{(1,4)}$，$Z=A^{(1,3)}$，$X=YAZ$，据引理3.1.8可知$X\in A\{1,\ 2\}$；另外，还有

$$AX=AYAZ=AZ,\ XA=YAZA=YA$$

注意到AZ，YA都是Hermite阵，所以AX，XA也都是Hermite阵。从而证明了$X\in A\{1,\ 2,\ 3,\ 4\}$。又因为$A\{1,\ 2,\ 3,\ 4\}=\{A^+\}$是单元素集合，故$X=A^+$。

定理3.1.4表明$A\in \mathbf{C}_r^{m\times n}$的$\{1\}$-逆有无穷多，且秩可取$[r,\ \min(m,\ n)]$中的任一整数。而$\{2\}$-逆的情况如何呢？从定义可知，$\operatorname{rank}(A^{(2)})\leqslant \operatorname{rank}(A)=r$。显然，零矩阵$O\in \mathbf{C}^{n\times m}$是$A$的一个$\{2\}$-逆，说明$A^{(2)}$的秩可达到0；另外，每个$A^{(1,2)}$是$A$的一个$\{2\}$-逆，且$\operatorname{rank}(A^{(1,2)})=\operatorname{rank}(A)=r$，说明$A^{(2)}$的秩可达到$r$；而一般情况下，有

定理3.1.13 设$A\in \mathbf{C}_r^{m\times n}$，$r>0$；则对任意指定的属于$[0,r]$的正整数$s$，一定存在$X_s\in A\{2\}$，满足

$$\operatorname{rank}(X_s)=s$$

证明 设$X_0\in A\{1,\ 2\}$，则有$\operatorname{rank}(X_0)=\operatorname{rank}(A)=r$。对$X_0$作满秩分解：

$$X_0=YZ,\qquad Y\in \mathbf{C}_r^{n\times r},\qquad Z\in \mathbf{C}_r^{r\times m}$$

则有

$$YZAYZ=X_0AX_0=X_0=YZ$$

据引理3.1.6，$Y^{(1)}=Y_L^{-1}$，$Z^{(1)}=Z_R^{-1}$，故上式两端左乘$Y^{(1)}$，右乘$Z^{(1)}$，便得到

$$ZAY=I_r$$

记Y_s为Y的前s列构成的子矩阵，Z_s为Z的前s行构成的子矩阵，则据$ZAY=I_r$很容易证明

$$Z_sAY_s=I_s$$

若令$X_s=Y_sZ_s\in \mathbf{C}^{n\times m}$，显然rank$(X_s)=s$，且有

$$X_sAX_s=Y_sZ_sAY_sZ_s=Y_sZ_s=X_s$$

第二节　M-P逆的几种显式表示

在第二章我们曾给出矩阵的几种分解式，现在我们用它们导出$M-P$逆的一些显式表示。

定理3.2.1(Macduffee，用秩分解构造A⁺)　设$A\in \mathbf{C}_r^{m\times n}$，且有满秩分解$A=FG$，$F\in \mathbf{C}_r^{m\times r}$，$G\in \mathbf{C}_r^{r\times n}$，则有

$$A^+=G^H(F^HAG^H)^{-1}F^H$$

证明　由于$F^HAG^H=(F^HF)(GG^H)$及F^HF，GG^H可逆故$(F^HAG^H)^{-1}$存在；令$X=G^H(F^HAG^H)^{-1}F^H$，很容易验证X满足Penrose的四个方程，故有$A^+=X$。

推论3.2.2　在定理3.2.1的假定下，有

$A^+=G^+F^+$，$G^+=G^H(GG^H)^{-1}$，$F^+=(F^HF)^{-1}F^H$

定理3.2.3　设$A\in \mathbf{C}_r^{m\times n}$，$r>0$，且有正交分解式

$$A=U\begin{bmatrix} R_r & 0 \\ 0 & 0 \end{bmatrix}V^H$$

其中U，V为酉阵，R_r是正线上三角阵。则有

$$A^+=V\begin{bmatrix} R_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}U^H$$

证明 代入Penrose四个方程直接验证便知。

推论3.2.4 设$A\in \mathbf{C}_r^{m\times n}(r>0)$有如下的奇异值分解

$$A=U\begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}V^H$$

其中 $S=\text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$，$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_r>0$ 为A的奇异值。则有

$$A^+=V\begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}U^H$$

下面考虑用半正定的Hermite矩阵$A^H A$的谱分解式给出A^+的显式表示。设$A^H A$的互异特征值为λ_1，λ_2，…，λ_k，则存在k个正交投影算子：$P_1, P_2, \cdots, P_k$，使得$A^H A=\sum_{i=1}^{k}\lambda_i P_i$，

不难验证有 $(A^H A)^+=\sum_{i=1}^{k}\lambda_i^+ P_i$，其中

$$\lambda_i^+=\begin{cases} 0, & \lambda_i=0 \\ \lambda_i^{-1}, & \lambda_i\neq 0 \end{cases}$$

定理3.2.5 设$A\in \mathbf{C}^{m\times n}$，$\lambda_1$，$\lambda_2$，…，$\lambda_k$为$A^H A$ 的互异特征值且$A^H A$具有谱分解式$A^H A=\sum_{i=1}^{k}\lambda_i P_i$。则有

$$A^+=(A^H A)^+A^H$$

证明 利用谱算子 P_i 的性质很容易验证上面的 A^+ 满足Penrose的四个方程，留作练习。

第三节 矩阵方程AXB=D的求解与广义逆的性质

本节首先用{1}-逆讨论矩阵方程$AXB=D$的求解问题，

然后论述各种弱逆的特性，也就是要分析为什么Penrose 用这样四个方程来定义A^+，并且证明Moore逆与Penrose逆是等价的。

定理3.3.1 设 $A\in \mathbf{C}^{m\times n}$，$B\in \mathbf{C}^{p\times q}$，$D\in^{m\times q}$ 。则矩阵方程$AXB=D$相容的充分必要条件是对每个$A^{(1)}$及$B^{(1)}$都有

$$AA^{(1)}DB^{(1)}B=D \qquad (*)$$

且这时的通解为

$$X=A^{(1)}DB^{(1)}+Y-A^{(1)}AYBB^{(1)} \qquad (**)$$

其中$Y\in \mathbf{C}^{n\times p}$是任意的。

证明 必要性：若X是方程$AXB=D$ 的一个解，则对每个$A^{(1)}A$及$B^{(1)}$都有

$$D=AXB=AA^{(1)}AXBB^{(1)}B=AA^{(1)}DB^{(1)}B$$

此即证明了(＊)成立。

充分性是显然的：若(＊) 成立，则取 $X=A^{(1)}DB^{(1)}$便是方程的解，从而说明 $AXB=D$ 是相容的。

将(＊＊)中的X代入方程$AXB=D$ 的左端并利用（＊）及{1}-逆的定义，可推出等于D,说明X是方程$AXB=D$的解。反过来，设X_0为$AXB=D$的任一解，则由于$AX_0B=D$，故可写为

$$X_0=A^{(1)}DB^{(1)}+X_0-A^{(1)}DB^{(1)}$$
$$=A^{(1)}DB^{(1)}+X_0-A^{(1)}AX_0BB^{(1)}$$

它相当于在（＊＊）中取$Y=X_0$，故有通解的形式。

由上面定理立即可得到

推论3.3.2 设$A\in \mathbf{C}^{m\times n}$，$b\in \mathbf{C}^m$，则线性方程组$Ax=b$ 相容的充分必要条件是对每个$A^{(1)}$都有

$$AA^{(1)}b=b$$

且这时方程$Ax=b$的通解为

$$x=A^{(1)}b+(I-A^{(1)}A)y$$

其中$y\in \mathbf{C}^n$是任意的。

推论3.3.3 设$A\in \mathbf{C}^{m\times n}$，$A^{(1)}\in A\{1\}$；则有

$$A\{1\}=\{A^{(1)}+Z-A^{(1)}AZAA^{(1)}\mid \forall Z\in \mathbf{C}^{n\times m}\}$$

证明 据定义

$$A\{1\}=\{X\in \mathbf{C}^{n\times m}\mid AXA=A\}$$

根据定理3.3.1可知，$AXA=A$的通解为

$$X=A^{(1)}AA^{(1)}+Y-A^{(1)}AYAA^{(1)},\quad \forall Y\in \mathbf{C}^{n\times m}$$

若令$Y=A^{(1)}+Z$，则上式又可写成

$$\begin{aligned}X&=A^{(1)}+Z+A^{(1)}AA^{(1)}-A^{(1)}AA^{(1)}+A^{(1)}AZAA^{(1)}\\&=A^{(1)}+Z-A^{(1)}AZAA^{(1)},\quad \forall Z\in \mathbf{C}^{n\times m}\end{aligned}$$

上述推论用某一个给定的$A^{(1)}$，便给出了集合$A\{1\}$的全部元素。

定理3.3.4 设$A\in \mathbf{C}^{m\times n}$，$G\in \mathbf{C}^{n\times m}$。则$G\in A\{1\}$的充分必要条件是：对所有使$Ax=b$相容的$b\in \mathbf{C}^{m}$，$x=Gb$皆为其解。

证明 充分性：设$a_j\in \mathbf{C}^{m}$为A的第j个列向量，则方程$Ax=a_j$肯定是相容的，并且Ga_j为其一解，即

$$AGa_j=a_j,\qquad j=1,2,\cdots,n$$

从而便有

$$\begin{aligned}AGA=AG(a_1a_2\cdots a_n)&=(AGa_1AGa_2\cdots AGa_n)\\&=(a_1a_2\cdots a_n)=A\end{aligned}$$

即证明了$G\in A\{1\}$。

必要性：根据推论3.3.2是显然的。

定理3.3.5(Penrose) 矩阵方程

$$AX=B,\qquad XD=E$$

具有公共解X的充分必要条件是：每个方程都有解，并且$AE=BD$。

证明 充分性：若方程$AX=B$，$XD=E$都相容，则应有$AA^{(1)}B=B$，$ED^{(1)}D=E$。那么，在$AX=B$的通解$X=A^{(1)}B+(I-A^{(1)}A)Y$中取$Y=ED^{(1)}$，便有$X=A^{(1)}B+ED^{(1)}-A^{(1)}AED^{(1)}$，这时很容易验证：$XD=A^{(1)}BD$

$+ED^{(1)}D - A^{(1)}AED^{(1)}D = A^{(1)}AE + E - A^{(1)}AE = E$

这说明$AX=B$与$XD=E$有公共解。

必要性： 若$AX=B$，$XD=E$，有公共解X，则应有$AXD=BD$，$AXD=AE$，从而有$AE=BD$。

定理3.3.6 集合$A\{1,3\}$是矩阵方程

$$AX = AA^{(1,3)}$$

的解集，其中$A^{(1,3)}$是$A\{1,3\}$中的任一元。

证明 当X_0是$AX=AA^{(1,3)}$的解时，即$AX_0=AA^{(1,3)}$，则有$AX_0A=AA^{(1,3)}A=A$，，说明$X_0\in A\{1\}$；另外，因$AA^{(1,3)}$是Hermite阵，故AX_0亦是，即$X_0\in A\{3\}$，从而证明了$X_0\in A\{1,3\}$。反之，若$X_0\in A\{1,3\}$，则有

$$\begin{aligned} AA^{(1,3)} &= AX_0AA^{(1,3)} = (AX_0)^H(AA^{(1,3)})^H \\ &= X_0^HA^H(A^{(1,3)})^HA^H = X_0^HA^H \\ &= (AX_0)^H = AX_0 \end{aligned}$$

这表明$A\{1,3\}$中的每个元皆为$AX=AA^{(1,3)}$的解。

根据定理3.3.1，方程$AX=AA^{(1,3)}$的通解为

$$X = A^{(1)}AA^{(1,3)} + Y - A^{(1)}AY$$

令$Y=Z+A^{(1,3)}$便得到

推论3.3.7 设$A\in \mathbf{C}^{m\times n}$，$A^{(1,3)}\in A\{1,3\}$。则

$$A\{1,3\} = \{A^{(1,3)} + (I - A^{(1,3)}A)Z \mid Z\in \mathbf{C}^{n\times m}\}$$

类似地推导过程可得到关于$A\{1,4\}$相应的结果：

定理3.3.8 集合$A\{1,4\}$为矩阵方程

$$XA = A^{(1,4)}A$$

的解集，其中$A^{(1,4)}$为$A\{1,4\}$中的任一元。

推论3.3.9 设$A\in \mathbf{C}^{m\times n}$，$A^{(1,4)}\in A\{1,4\}$。则

$$A\{1,4\} = \{A^{(1,4)} + Z(I - AA^{(1,4)}) \mid Z\in \mathbf{C}^{n\times m}\}$$

請读者按上述的推导方法，给出集合$A\{1,3,4\}$，$A\{1,2,3\}$$A\{1,2,4\}$的表达式。

前面简单地讨论了满足四个Penrose方程中的某几个方程的弱

广义逆的特征，真正地揭示出它们的性质，还要在下面关于矛盾方程最小二乘解的讨论中。我们先讨论A^+的性质。

引理3.3.10 设$A\in \mathbf{C}^{m\times n}$，$A^+$ 为$M-P$广义逆，即 A^+ 为 $A\{1,2,3,4\}$中的唯一元。则有

1. $\mathscr{R}(A)=\mathscr{R}(AA^+)\subset \mathbf{C}^m$；
2. $\mathscr{R}(A^+A)=\mathscr{R}(A^+)=\mathscr{R}(A^H)\subset \mathbf{C}^n$；
3. $\mathscr{N}(A)=\mathscr{N}(A^+A)\subset \mathbf{C}^n$；
4. $\mathscr{N}(AA^+)=\mathscr{N}(A^+)=\mathscr{N}(A^H)\subset \mathbf{C}^m$。

证明 1. $\mathscr{R}(AA^+)\subset\mathscr{R}(A)$是显然的；$\forall x\in\mathscr{R}(A)$，则存在$y\in\mathbf{C}^n$，使得$x=Ay$，故有

$$x=Ay=AA^+Ay=(AA^+)(Ay)\in\mathscr{R}(AA^+)$$

这表明$\mathscr{R}(A)\subset\mathscr{R}(AA^+)$。从而证明了1是成立的。

2中的第一等式类似于1，用同样的方法可证。下面证明其中的第二个等式：$\mathscr{R}(A^+)=\mathscr{R}(A^H)$。设$x\in\mathscr{R}(A^H)$，则存在$y\in\mathbf{C}^m$，使得$x=A^H y$，故

$$\begin{aligned}x&=A^H y=(AA^+A)^H y=(A^+A)^H A^H y\\&=A^+(AA^H y)\in\mathscr{R}(A^+)\end{aligned}$$

这表明$\mathscr{R}(A^H)\subset\mathscr{R}(A^+)$；反之，若$x\in\mathscr{R}(A^+)$，则存在$y\in\mathbf{C}^m$，使得$x=A^+y$，故

$$\begin{aligned}x&=A^+y=A^+AA^+y=(A^+A)^H A^+y\\&=A^H((A^+)^H A^+y)\in\mathscr{R}(A^H)\end{aligned}$$

这表明$\mathscr{R}(A^+)\subset\mathscr{R}(A^H)$。从而证明了2是成立的。

3与4的证明，留作练习。

引理3.3.11 设$A\in\mathbf{C}^{m\times n}$，$A^+$为$M-P$广义逆。则有

1. $\mathscr{R}(A)^\perp=\mathscr{N}(A^H)=\mathscr{N}(A^+)$；
2. $\mathscr{R}(A^H)^\perp=\mathscr{N}(A)$。

证明 1.设$(\cdot|\cdot)$表示酉空间$\mathbf{C}^m$或 $\mathbf{C}^n$ 中的标准内积，则对于$A\in\mathbf{C}^{m\times n}$及任意的$x\in\mathbf{C}^m$及$y\in\mathbf{C}^n$，都有

$$(A^H x|y)=y^H A^H x=(Ay)^H x=(x|Ay)$$

设$x\in\mathscr{N}(A^H)$，则对一切$y\in\mathbf{C}^n$ 上式最左端等于零，故据上式最右端可知$x\perp Ay$，即$x\in\mathscr{R}(A)^\perp$。这表明$\mathscr{N}(A^H)\subset\mathscr{R}(A)^\perp$。反之，若$x\in\mathscr{R}(A)^\perp$，则对一切$y\in\mathbf{C}^n$，上式的最右端都为零，故据最左端对一切$y\in\mathbf{C}^n$都有$(A^H x|y)=0$，特别取$y=A^H x$ 亦成立。从而据内积公理的第一条，便得$A^H x=0$，这意味着$x\in\mathscr{N}(A^H)$，从而证明了$\mathscr{R}(A)^\perp\subset\mathscr{N}(A^H)$。综合起来，便证明了$\mathscr{R}(A)^\perp=\mathscr{N}(A^H)$。

对于2，亦可类似的证明。不过，从下面的等价关系也很容易推出：

$$x\in\mathscr{N}(A)\Longleftrightarrow Ax=0\Longleftrightarrow x^H A^H=0\Longleftrightarrow x\perp\mathscr{R}(A^H)$$
$$\Longleftrightarrow x\in\mathscr{R}(A^H)^\perp。$$

引理3.3.12 设$\mathbf{C}^n=L\dot{+}M$。则当$M=L^\perp$时，$P_{L,M}$具有Hermite性，即$P_{L,M}^H=P_{L,M}$。反之亦然。

证明 据引理3.3.11有$\mathscr{N}(A)=\mathscr{R}(A^H)^\perp$，故而有

$$\mathscr{R}(P_{L,M}^H)=\mathscr{N}(P_{L,M})^\perp=M^\perp,$$
$$\mathscr{N}(P_{L,M}^H)=\mathscr{R}(P_{L,M})^\perp=L^\perp,$$

这表明

$$P_{L,M}^H=P_{M^\perp,L^\perp}=P_{L,M}$$

定理3.3.13 设$A\in\mathbf{C}^{m\times n}$，$A^+$为$A$的$M-P$ 广义逆。则下列条件是等价的：

1. $AA^+A=A$，$A^+AA^+=A^+$，$(AA^+)^H=AA^+$，$(A^+A)^H=A^+A$；
2. $A^H AA^+=A^H$，$(A^+)^H A^+A=(A^+)^H$；
3. $AA^+=P_{\mathscr{R}(A),\mathscr{N}(A^H)}$
   $A^+A=P_{\mathscr{R}(A^H),\mathscr{N}(A)}$

证明 1⟶2是显然的；现证2⟶1：首先由$A^H AA^+=A^H$可知$A=(AA^+)^H A$，从而便有

$$(AA^+)^H=(A^+)^H A^H=(A^+)^H A^H AA^+$$

$$=(AA^+)^H AA^+=AA^+$$

同样由$(A^+)^H A^+A=(A^+)^H$可推出$(A^+A)^H=A^+A$。另外还可推出

$$A=(A^H AA^+)^H=(AA^+)^H A=AA^+A$$

$$A^+=((A^+)^H A^+A)^H=(A^+A)^H A^+=A^+AA^+$$

1⟶3：因为AA^+，A^+A都是幂等的Hermite 矩阵，故都是正交投影算子：

$$AA^+=P_{\mathscr{R}(AA^+),\ \mathscr{R}(AA^+)^\perp}$$

$$A^+A=P_{\mathscr{R}(A^+A),\ \mathscr{R}(A^+A)^\perp}$$

但据引理3.3.10及3.3.11知，$\mathscr{R}(AA^+)=\mathscr{R}(A)$，$\mathscr{R}(AA^+)^\perp=\mathscr{N}(A^H)$，$\mathscr{R}(A^+A)=\mathscr{R}(A^H)$，$\mathscr{R}(A^+A)^\perp=\mathscr{N}(A)$，故而有

$$AA^+=P_{\mathscr{R}(A),\ \mathscr{N}(A^H)},\quad A^+A=P_{\mathscr{R}(A^H),\ \mathscr{N}(A)}。$$

反过来，据引理3.3.11知：

$$\mathbf{C}^m=\mathscr{R}(A)\dot{+}\mathscr{N}(A^H),\quad \mathbf{C}^n=\mathscr{R}(A^H)\dot{+}\mathscr{N}(A)$$

都是正交直和分解，故投影算子$AA^+=P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}$及$A^+A=P_{\mathscr{R}(A^H),\ \mathscr{N}(A)}$都是幂等的，即可推出Penrose的前两个方程；而据引理 3.3.12 知，投影算子 AA^+，A^+A 也都是Hermite的，从而推出Penrose的后两个方程，即完成 3⟶1 的证明。

这个定理证明了Moore与Penrose的广义逆是等价的，都定义了唯一的A^+。

第四节 广义逆与矛盾方程組

Ax＝b的求解

设$A\in\mathbf{C}^{m\times n}$，$b\in\mathbf{C}^m$。线性方程组$Ax=b$当且仅当$b\in\mathscr{R}(A)$是

相容的。上一节也用{1}-逆给出了方程组$Ax=b$相容的另一个充分必要条件及通解的表达式。当$b\notin\mathscr{R}(A)$时，不存在任何$x\in\mathbf{C}^n$使得$Ax=b$成立，所以这时称之为矛盾方程。给定了A及b，既然对一切$x\in\mathbf{C}^n$，残差向量

$$r=b-Ax$$

都不等于零，因此就希望求一个$x_0\in\mathbf{C}^n$使得

$$\|b-Ax_0\|=\min\{\|b-Ax\|\mid x\in\mathbf{C}^n\}$$

其中$\|\cdot\|$为酉空间$\mathbf{C}^m$中的（由内积诱导出的）范数。这个x_0就是通常所说的矛盾方程的最小二乘解。在上述意义下，当$b\in\mathscr{R}(A)$时，x_0就是相容性方程的解。因此，上述最优化问题的求解方法适用于对方程$Ax=b$的一般讨论。如果我们令$y_0=Ax_0$，$y=Ax$，则上述优化问题就是求b在$\mathscr{R}(A)$上的最佳逼近y_0；据第一章讲过的：y_0为b在$\mathscr{R}(A)$上的最佳逼近的充要条件是$b-y_0\perp\mathscr{R}(A)$；即由此可推出最佳逼近的法方程

$$A^H Ax_0=A^H b$$

这也是矛盾方程$Ax=b$的最小二乘解x_0应该满足的方程，因此也叫做矛盾方程$Ax=b$的法方程。下面的定理将给出$A\{1,3\}$与矛盾方程$Ax=b$的最小二乘解之间的关系。

定理3.4.1 设$A\in\mathbf{C}^{m\times n}$，$b\in\mathbf{C}^m$，任取$A^{(1,3)}\in A\{1,3\}$，则$x_0=A^{(1,3)}b$是$Ax=b$的一个最小二乘解。反之，如果$G\in\mathbf{C}^{n\times m}$对一切$b\in\mathbf{C}^m$都使得$x_0=Gb$是$Ax=b$的一个最小二乘解，则$G\in A\{1,3\}$。

证明 当$G\in A\{1,3\}$时，有$AGA=A$及$(AG)^H=AG$，故$A^H=A^H(AG)^H=A^H AG$，令$x_0=Gb$，便有

$$A^H Ax_0=A^H AGb=A^H b$$

说明$x_0=Gb$是满足$Ax=b$的法方程，故是$Ax=b$的一个最小二乘解。

反之，若对每个$b\in\mathbf{C}^m$，$x_0=Gb$都满足法方程，即$A^H AGb=A^H b$，亦即$(A^H AG-A^H)b=0$；把$A^H AG-A^H$视为从

$\mathbf{C}^m$到$\mathbf{C}^n$的线性映射，由于$b\in\mathbf{C}^m$的任意性，知$A^H AG-A^H=0$，即$A^H=A^H AG$。亦即有$A=(AG)^H A$，两端右乘G，便有$AG=(AG)^H AG$；这表明AG是一个Hermite阵，即$G\in A\{3\}$。从而据$A=(AG)^H A=AGA$又知$G\in A\{1\}$，故$G\in A\{1, 3\}$。

由$AA^{(1,3)}$是一个幂等的Hermite阵，故它是一个正交投影算子：$AA^{(1,3)}=P_{\mathscr{R}(AA^{(1,3)}),\mathscr{R}(AA^{(1,3)})^{\perp}}$。但由于$\mathscr{R}(AA^{(1,3)})=\mathscr{R}(A)$，故$\mathscr{R}(AA^{(1,3)})^{\perp}=\mathscr{R}(A)^{\perp}\mathscr{N}(A)^H$，从而$AA^{(1,3)}=P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}$。这样便给出：

推论3.4.2　向量$x_0\in\mathbf{C}^n$为方程$Ax=b$的最小二乘解，当且仅当有

$$Ax_0=AA^{(1,3)}b=P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}b$$

并且最小二乘解的通式为

$$x_0=A^{(1,3)}b+(I_n-A^{(1,3)}A)y$$

其中$A^{(1,3)}\in A\{1, 3\}$，$y\in\mathbf{C}^n$都是任取的。

从通式可以看出，只有A是列满秩时，最小二乘解才是唯一的，且为$x_0=(A^H A)^{-1}A^H b$。否则，便有无穷多个最小二乘解。

引理3.4.3　设$A\in\mathbf{C}^{m\times n}$，$b\in\mathscr{R}(A)$。则相容性方程

$$Ax=b$$

在$\mathscr{R}(A^H)$上有唯一的解x_0，且它是$Ax=b$的所有解中具有最小范数者。

证明　相容性方程$Ax=b$在$\mathscr{R}(A^H)$上有解是显然的：因为$\mathbf{C}^n=\mathscr{R}(A^H)\dot{+}\mathscr{R}(A^H)^{\perp}=\mathscr{R}(A^H)\dot{+}\mathscr{N}(A)$，若$x\in\mathbf{C}^n$是$Ax=b$的一解，则它有唯一的分解

$$x=x_0+y,\ x_0\in\mathscr{R}(A^H),\ y\in\mathscr{N}(A)$$

从而有

$$b=Ax=A(x_0+y)=Ax_0$$

这表明$x_0\in\mathscr{R}(A^H)$是$Ax=b$的一个解。

设$u,v\in\mathscr{R}(A^H)$是$Ax=b$的两个解,则首先有$u-v\in\mathscr{R}(A^H)$,进而有$A(u-v)=Au-Av=b-b=0$，这表明$u-v\in\mathscr{N}(A)$，但因$\mathscr{R}(A^H)\cap\mathscr{N}(A)=\{0\}$，故$u-v=0$，从而证明了$Ax=b$在$\mathscr{R}(A^H)$上有唯一的解。

设$Ax=b$的通解为

$$x=x_0+y,\ x_0\in\mathscr{R}(A^H),\ y\in\mathscr{N}(A)$$

则由于$x_0\perp y$，故据勾股定理有

$$\|x\|^2=\|x_0\|^2+\|y\|^2$$

这表明，除非$x=x_0$，恒有$\|x\|>\|x_0\|$。从而证明了 $Ax=b$在$\mathscr{R}(A^H)$上的唯一解x_0是它的最小范数解。

下面的定理把A的$\{1,4\}$-逆与$Ax=b$的最小范数解联系起来。

定理3.4.4 设$A\in\mathbf{C}^{m\times n}$，$b\in\mathbf{C}^m$，$A^{(1,4)}\in A\{1,4\}$。假若方程$Ax=b$有解（即$b\in\mathscr{R}(A)$），则$x_0=A^{(1,4)}b$ 是其最小范数解。反之，若$G\in\mathbf{C}^{n\times m}$对一切$b\in\mathscr{R}(A)$都使得$x=Gb$是$Ax=b$的最小范数解，则一定有$G\in A\{1,4\}$。

证明 若方程相容，据推论 3.3.2 及引理 3.1.7 知$x_0=A^{(1,4)}b$是$Ax=b$的解且$x_0\in\mathscr{R}(A^H)$，从而据引理 3.4.3 知它是$Ax=b$的唯一的最小范数解。

反之，若对一切$b\in\mathscr{R}(A)$，$x=Gb$都是方程$Ax=b$的最小范数解，则据前面的证明：对指定的$b\in\mathscr{R}(A)$，任取$A^{(1,4)}\in A\{1,4\}$，$A^{(1,4)}b$ 是 $Ax=b$ 唯一的最小范数解，故而有 $Gb=A^{(1,4)}b$。现在让b遍取A的各列，则得到G应满足的方程

$$GA=A^{(1,4)}A$$

从而据定理3.3.8知$G\in A\{1,4\}$。

至于$Ax=b$的唯一的最小范数最小二乘解，是与M-P广义逆有联系的。

定理3.4.5(Penrose) 设$A\in\mathbf{C}^{m\times n}$，$b\in\mathbf{C}^m$。则方程 $Ax=b$的所有解或是所有最小二乘解中，A^+b是其中的最小范数解。反

之，如果$G\in\mathbf{C}^{n\times m}$能够对一切$b\in\mathbf{C}^m$，都使得$Gb$是$Ax=b$的最小范数最小二乘解，则一定有$G=A^+$。

证明 由推论3.4.2可知，$Ax=b$的最小二乘解即为方程$Ax=AA^{(1,3)}b$的解；因此$Ax=b$的最小范数最小二乘解即为方程$Ax=AA^{(1,3)}b$的最小范数解。从而据定理3.4.4可知，这个解为

$$x=A^{(1,4)}AA^{(1,3)}b$$

而据定理3.1.12，$A^{(1,4)}AA^{(1,3)}=A^+$，故$x=A^+b$。

反之，如果$G\in\mathbf{C}^{n\times m}$对一切$b\in\mathbf{C}^m$都使得$Gb$为$Ax=b$的最小范数最小二乘解，则据前面的证明知，对一切$b\in\mathbf{C}^m$都有$Gb=A^+b$，让$b$遍取单位矩阵$I_m$的各列，便证得$G=A^+$。

这个定理和以前的定理表明了：给定方程$Ax=b$，只要计算出A^+，A^+b便给出了方程的各种意义下的解。即当$Ax=b$相容时，A^+b要末是它的唯一解，要末是唯一的最小范数解；当$Ax=b$不相容时，A^+b要末是它的唯一最小二乘解，要末是唯一的最小范数最小二乘解。所以说，$M-P$广义逆A^+把方程组$Ax=b$的求解问题，从理论上和方法上都完满地解决了。所以具体可行的计算A^+的方案，是下一步要解决的主要问题。

第五节 几种計算$\mathbf{A}^+$的方法

我们首先给出A^+的Lagrange-Sylvester公式及Neumann展式，然后再简单地介绍几种计算方法。

设A^HA的非零互异特征值为σ_1^2，σ_2^2，…，σ_k^2，而A^HA的最小多项式为

$$g(\lambda)=\prod_{i=1}^{k}(\lambda-\sigma_i^2)$$

并令

$$g_i(\lambda)=\prod_{j\neq i}(\lambda-\sigma_j^2)$$

又设$A^H A$的谱分解式为

$$A^H A=\sum_{i=1}^{k}\sigma_i^2 P_i$$

则对λ的任一多项式$\varphi(\lambda)$都有

$$\varphi(A^H A)=\sum_{i=1}^{k}\varphi(\sigma_i^2)P_i$$

由于当$j\neq i$时有$g_i(\sigma_j^2)=0$，故而有

$$g_i(A^H A)=g_i(\sigma_i^2)P_i$$

由上式可以得到

$$P_i=g_i(A^H A)/g_i(\sigma_i^2)$$

$$=\prod_{j\neq i}(A^H A-\sigma_j^2 I)/\prod_{j\neq i}(\sigma_i^2-\sigma_j^2)$$

又因为

$$(A^H A)^+=\sum_{i=1}^{k}\sigma_i^{-2}{}_iP$$

故而有

$$A^+=(A^H A)^+A^H$$

$$=\sum_{i=1}^{k}\sigma_i^{-2}\frac{\prod_{j\neq i}(A^H A-\sigma_j^2 I)}{\prod_{j\neq i}(\sigma_i^2-\sigma_j^2)}A^H$$

这就是所谓的A^+的Lagrange-Sylvester公式。

设$A\in\mathbf{C}_r^{m\times n}$的奇异值分解为$A=UDV^H$，其中

$$D=\begin{bmatrix} S & 0\\ 0 & 0\end{bmatrix}$$

则有

$$A^+=V\begin{bmatrix} S^{-1} & 0\\ 0 & 0\end{bmatrix}U^H$$

其中$S^{-1}=\mathrm{diag}(\sigma_1^{-1}I_{r_1},\sigma_2^{-1}I_{r_2},\cdots,\sigma_k^{-1}I_{r_k})$，$\sum_{i=1}^{k}r_i=r$。

定理3.5.1（Nenmann） 设$A\in\mathbf{C}_r^{m\times n}$，$\sigma_1,\sigma_2\cdots,\sigma_n$为其$n$个奇异值，又令$c=\max(\sigma_1^2,\sigma_2^2\cdots,\sigma_n^2)$，$0<\alpha<2/c$。则Neumann展式$\alpha\sum_{k=0}^{\infty}(I-\alpha A^H A)^k A^H$收敛于$A^+$。

证明 因为$A^H A=VD^H DV^H$，其中

$$D^H D=\begin{bmatrix}S^2 & 0\\ 0 & 0\end{bmatrix}$$

故有

$$(I-\alpha A^H A)^k=V(I-\alpha D^H D)^k V^H$$
$$\alpha(I-\alpha A^H A)^k A^H=V\alpha(I-\alpha D^H D)^k D^H U^H$$

很显然$W=\alpha\sum_k(I-\alpha D^H D)^k D^H$为对角阵，并且若$D^H$的对角元为零时，$W$的相应的对角元也为零，而$D^H$的对角元为$\sigma_i$时，$W$的相应的对角元为

$$\alpha\sum_k(1-\alpha\sigma_i^2)^k\sigma_i=\sigma_i^{-1}$$

这表明$W=D^+$，从而证明了定理3.5.1。

值得注意的是，当$A\in\mathbf{C}_n^{n\times n}$时，$\alpha\sum_{k=0}^{\infty}(I-\alpha A^H A)^k A^H$就是$A^{-1}$的Neumann展式。

推论3.5.2 在定理3.5.1的条件下，还有

$$A^+=\alpha\sum_{k=0}^{\infty}A^H(I-\alpha AA^H)^k$$

证明 根据定理3.5.1有

$$(A^H)^+=\alpha\sum_{k=0}^{\infty}(I-\alpha AA^H)^k A$$

故有

$$A^+=[(A^H)^+]^H=\alpha\sum_{k=0}^{\infty}A^H(I-\alpha AA^H)^k$$

定理3.5.3（Neumann-Eujer展示） 在定理3.5.1的条件下，设

$$A_p^{+}=\alpha\left\{[I+(I-\alpha A^H A)]\prod_{k=1}^{p-1}[I+(I-\alpha A^H A)^{2^k}]\right\}A^H$$

则

$$\|A^{+}-A_p^{+}\|\leqslant\max_i\frac{(1-\alpha\sigma_i^{2})^{2^p}}{\sigma_i}$$

其中矩阵M的范数定义为

$$\|M\|=\max\{\sqrt{\lambda_i}\mid\lambda_i\text{为}M^H M\text{的特征值}\}$$

证明 根据Euler恒等式

$$(1+x)\prod_{k=1}^{p-1}(1+x^{2^k})=\prod_{k=0}^{2^p-1}x^k$$

我们考察

$$A^{+}-A_p^{+}=\alpha\sum_{k=2^p}^{\infty}(I-\alpha A^H A)^k A^H$$

$$=\alpha V\sum_{k=2^p}^{\infty}(I-\alpha D^H D)^k D^H U^H$$

则有

$$\|A^{+}-A_p^{+}\|\leqslant\alpha\|V\|\left\|\sum_{k=2^p}^{\infty}(I-\alpha D^H D)^k D^H\right\|\|U^H\|$$

$$=\max_i\frac{(1-\alpha\sigma_i^2)^{2^p}}{\sigma_i}$$

推论3.5.4 在定理3.5.1的条件下，设

$$A_p^{+}=\alpha A^H\left\{[I+(I-\alpha AA^H)]\prod_{k=1}^{p-1}[I+(I-\alpha AA^H)^{2^k}]\right\}$$

则有

$$\|A^{+}-A_p^{+}\|\leqslant\max_i\frac{(1-\alpha\sigma_i^{2})^{2^p}}{\sigma_i}$$

根据A^+的 Neumann 展式，可以考虑计算A^+的迭代方法，目的是给出一个收敛于A^+的矩阵序列$\{X_k\}$。迭代过程的收敛速度取决于相应的残差序列：

$$R_k = P_{\mathscr{R}(A),\mathscr{N}(A^H)} - AX_k, \qquad k=0,1,2,\cdots$$

当$X_k \longrightarrow A^+$时，$R_k \longrightarrow 0$。对于这一迭代过程，要求存在一个正整数p，使得对满足 $\|AB\| \leqslant \|A\|\,\|B\|$ 的任一种范数都有

$$\|R_{k+1}\| \leqslant C\|R_k\|^p, \quad k=0,1,2,\cdots$$

设想递推关系为

$$X_{k+1} = X_k + C_k R_k, \qquad k=0,1,2,\cdots \qquad (*)$$

其中$\{C_k\}$是某一适当的矩阵序列，X_0为待定的初值。用上式来计算A^+碰到的第一个困难是要给出R_k，而计算R_k又需要计算$P_{\mathscr{R}(A),\mathscr{N}(A^H)}$，它与计算$A^+$几乎有相同的工作量，所以递推关系式（$*$）看来是不能使用的。为克服这一困难，可通过选择适当的C_k来解决。若取

$$C_k = C_k P_{\mathscr{R}(A),\mathscr{N}(A^H)}, \qquad k=0,1,\cdots$$

则有

$$\begin{aligned} C_k R_k &= C_k\,(P_{\mathscr{R}(A),\mathscr{N}(A^H)} - AX_k) \\ &= C_k(I - AX_k) \end{aligned}$$

因而式($*$)可改写为

$$X_{k+1} = X_k + C_k T_k, \qquad k=0,1,\cdots \quad (**)$$

其中

$$T_k = I - AX_k, \qquad k=0,1,\cdots$$

若$A\in \mathbf{C}^{m\times n}$，由($*$)及($**$)给出的迭代算法适用于$m\leqslant n$的情形，因为这时的$R_k$与$T_k$都是$m\times m$阶方阵。可是，在$m>n$的情况下可使用其对偶的形式：

$$\tilde{X}_{k+1} = \tilde{X}_k + \tilde{R}_k \tilde{C}_k \qquad (*)'$$

其中

$$\tilde{R}_k = P_{\mathscr{R}(A^H),\ \mathscr{N}(A)} - \tilde{X}_k A$$

而$\tilde{C}_k$，$k=0$，1，…，为适当选取的序列，且满足

$$\tilde{C}_k = P_{\mathscr{R}(A^H),\ \mathscr{N}(A)}\tilde{C}_k,\quad k=0,\ 1,\ \cdots$$

从而可将(＊)′改写成

$$\tilde{X}_{k+1} = \tilde{X}_k + \tilde{T}_k\tilde{C}_k,\quad k=0,\ 1,\ \cdots \qquad (**)'$$

式中

$$\tilde{T}_k = I - \tilde{X}_k A,\quad k=0,\ 1,\ \cdots$$

事实上，$\tilde{T}_k$是$n\times n$的矩阵，因而当$m>n$时，这种迭代算法比前一种要好。下面仅就(＊)或(＊＊)给出的迭代序列进行讨论，不过结论也适用于(＊)′或(＊＊)′所给出的迭代序列。

定理3.5.5 设$0\neq A\in\mathbf{C}^{m\times n}$，初值$X_0$及残差$R_0$分别满足$X_0=A^H B_0 A^H$与$\rho(R_0)<1$，$B_0\in\mathbf{C}^{m\times n}$为某一矩阵，$\rho(R_0)$表示$R_0$的谱半径，即$\max\limits_{1\leqslant i\leqslant m}|\lambda_i(R_0)|$。则序列

$$\begin{aligned} X_{k+1} &= X_k + X_0 T_k \qquad\qquad (***)\\ &= X_k + X_0(I - AX_k),\ k=0,\ 1,\ \cdots \end{aligned}$$

当$k\longrightarrow\infty$时收敛于A^+,相应的残差对任意相容范数（即$\|AB\|\leqslant\|A\|\ \|B\|$被满足的范数）满足

$$\|R_{k+1}\|\leqslant\|R_0\|\ \|R_k\|,\quad k=0,\ 1,\cdots$$

证明 上面的计算序列是在(＊＊)中取 $C_k=X_0$ ($k=0$，1，…)而得到的。据$X_0=A^H B_0 A^H$可知

$$\begin{aligned} X_0 P_{\mathscr{R}(A),\ \mathscr{N}(A^H)} &= A^H B_0 A^H AA^+\\ &= A^H B_0(AA^+A)^H = A^H B_0 A^H = X_0 \end{aligned}$$

即满足条件 $C_k = C_k P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}$，故可将(＊＊)改写为

$$\begin{aligned} X_{k+1} &= X_k + X_0 R_k\\ &= X_k + X_0(P_{\mathscr{R}(A),\ \mathscr{N}(A^H)} - AX_k)\\ &\qquad k=0,\ 1,\ \cdots \end{aligned}$$

由此便求得残差

$$
\begin{aligned}
R_{k+1} &= P_{\mathscr{R}(A),\ \mathscr{N}(A^H)} - AX_{k+1} \\
&= P_{\mathscr{R}(A),\ \mathscr{N}(A^H)} - AX_k - AX_0R_k \\
&= R_k - AX_0R_k \\
&= P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}R_k - AX_0R_k \\
&= R_0R_k \qquad\qquad k=0,1,\cdots
\end{aligned}
$$

显然可推出

$$R_{k+1} = R_0R_k = \cdots = R_0^{k+2}$$

从而对于任意的相容矩阵范数有

$$\| R_{k+1} \| \leqslant \| R_0 \|^{k+2}$$

由假设条件$\rho(R_0)<1$，可知$\| R_{k+1} \| \longrightarrow 0$，即

$$P_{\mathscr{R}(A),\ \mathscr{N}(A^H)} - AX_k \longrightarrow 0 \qquad (k\longrightarrow\infty)$$

下面再证(* *)中的序列是收敛的，将(* *)改写成

$$X_{k+1} = X_k + X_0R_k$$

由式$R_{k+1} = R_0{}^{k+2}$可得

$$
\begin{aligned}
X_{k+1} &= X_k + X_0R_0^{k+1} \\
&= X_{k-1} + X_0R_0{}^k + X_0R_0^{k+1} \\
&= X_0(I + R_0 + \cdots + R_0{}^{k+1})
\end{aligned}
$$

由于$\rho(R_0)<1$,故序列X_{k+1}收敛于某个X。留下来的是要证明$X = A^+$。前面已证明$AX_k \longrightarrow P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}$，亦即$AX = P_{\mathscr{R}(A),\ \mathscr{N}(A^H)}$。因此便知$X\in A\{1,3\}$。根据$X_0 = A^HB_0A^H$及$X_{k+1} = X_k + X_0(I - AX_k)$可推得$X_k = A^HB_kA^H$，$B_k\in \mathbf{C}^{m\times n}$，从而有

$$X = \lim_{k\to\infty} X_k = A^HBA^H,\ B\in\mathbf{C}^{m\times n}。$$

这样便很容易推得

$$
\begin{aligned}
A^HAX &= A^H(AX)^H = (AXA)^H = A^H \\
X^HXA &= AB^HAXA = AB^HA = X^H
\end{aligned}
$$

因而说明X满足定理3.3.13中的2,故$X = A^+$。

上述定理给出的迭代算法的收敛速度是一阶的，下面给出一个$p(\geqslant 2)$阶的迭代算法。

定理3.5.6 在定理3.5.5的条件下，对任一整数$p \geqslant 2$，当$k \longrightarrow \infty$时序列

$$\begin{aligned} X_{k+1} &= X_k(I + T_k + T_k^2 + \cdots + T_k^{p-1}) \\ &= X_k[I + (I - AX_k) + (I - AX_k)^2 + \cdots \\ &\quad + (I - AX_k)^{p-1}] \end{aligned}$$

收敛于A^+，并且相应的残差序列满足

$$\| R_{k+1} \| \leqslant \| R_k \|^p \qquad k = 0, 1, \cdots$$

证明 只要在(* *)中取

$$C_k = X_k(I + T_k + T_k^2 + \cdots + T_k^{p-2})$$

便可得到本定理给出的迭代序列。据初值条件$X_0 = A^H B_0 A^H$及迭代关系式可知$X_k = A^H B_k A^H$，$B_k \in \mathbf{C}^{m \times n}$，因而据$C_k$的表达式知给出的序列满足

$$C_k = C_k P_{\mathscr{R}(A), \mathscr{N}(A^H)}$$

类似于定理3.5.5的推导，可得

$$X_{k+1} = X_k(I + R_k + R_k^2 + \cdots + R_k^{p-1})$$

从而有

$$\begin{aligned} R_{k+1} &= P_{\mathscr{R}(A), \mathscr{N}(A^H)} - AX_{k+1} \\ &= P_{\mathscr{R}(A), \mathscr{N}(A^H)} - AX_k(I + R_k + \cdots + R_k^{p-1}) \\ &= R_k - AX_k(R_k + R_k^2 + \cdots + R_k^{p-1}) \end{aligned}$$

由于对所有的$j = 1, 2, \cdots, p-1$都有

$$\begin{aligned} R_k^j - AX_k R_k^j &= P_{\mathscr{R}(A), \mathscr{N}(A^H)} R_k^j - AX_k R_k^j \\ &= R_k \cdot R_k^j = R_k^{j+1} \end{aligned}$$

因此前式又可写为

$$R_{k+1} = R_k^p$$

从而有$\| R_{k+1} \| \leqslant \| R_k \|^p$。定理其余部分的证明与定理3.5.5相同，故略。

前面二种求解A^+的迭代算法有如下关系：

定理3.5.7 设 $0\neq A\in\mathbf{C}^{m\times n}$，由定理3.5.5给出的迭代序列$\{X_k\}$与定理3.5.6 给出的迭代序列$\{\tilde{X}_j\}$在同样的初始条件$X_0$之下，虽收敛于同一极限$A^+$，但是在有限次逼近之间存在如下关系：

$$\tilde{X}_0=X_0$$

$$\tilde{X}_{j+1}=\tilde{X}_j(I+\tilde{T}_j+\tilde{T}_j^2+\cdots+\tilde{T}_j^{p-1}) \qquad (***)'$$

$$\tilde{T}_j=I-A\tilde{X}_j, \qquad j=0,1,\cdots$$

$$\tilde{X}_j=X_{p^j-1}, \qquad j=0,1,\cdots$$

证明 对j采取归纳法证明。当$j=0$时，关系式$\tilde{X}_j=X_{p^j-1}$显然成立成。设$\tilde{X}_j=X_{p^j-1}$成立，来证明$\tilde{X}_{j+1}=X_{p^{j+1}-1}$也成立。由于

$$X_k=X_0(I+R_0+R_0^2+\cdots+R_0^k)$$

再据归纳假设$\tilde{X}_j=X_{p^j-1}$，便可推得

$$\tilde{X}_j=X_0(I+R_0+R_0^2+\cdots+R_0^{p^j-1})$$

从而定理3.5.6中的迭代序列可写成

$$\tilde{X}_{j+1}=\tilde{X}_j(I+\tilde{R}_j+\tilde{R}_j^2+\cdots+\tilde{R}_j^{p-1})$$

由此即得

$$\begin{aligned}\tilde{R}_j&=P_{\mathscr{R}(A),\mathscr{N}(A^H)}-A\tilde{X}_j\\&=P_{\mathscr{R}(A),\mathscr{N}(A^H)}-AX_{p^j-1}\\&=R_{p^j-1}=R_0^{p^j}\end{aligned}$$

最后便有

$$\begin{aligned}\tilde{X}_{j+1}&=\tilde{X}_j(I+R_0^{p^j}+R_0^{2p^j}+\cdots+R_0^{(p-1)p^j})\\&=X_0(I+R_0+\cdots+R_0^{p^j-1})(I+R_0^{p^j}+\cdots\\&\quad+R_0^{(p-1)p^j})\\&=X_0(I+R_0+\cdots+R_0^{p^{j+1}-1})=X_{p^{j+1}-1}\end{aligned}$$

上述定理表明，用$(***)'$经j次迭代所得到的A^+的近似表达$\bar{X}_j$，用$(***)$的算法经p^j-1次才能得到。然而也不难发现，$(***)'$的每次迭代所需要的计算量比$(***)$要多。所以在实际应用中就要考虑选择适当的阶次p，使计算量最小。

下面讨论广义逆的另一种计算方法，也就是Greville方法，它是一种有限次的递推算法。为此，首先引入下面的符号：A_k为分块矩阵

$$A_k=[A_{k-1}a_k],\quad k=2,3,\cdots,n$$

其中a_k为A的第k列，且令

$$d_k=A_{k-1}^+a_k,\quad k=2,3,\cdots,n$$

$$\begin{aligned}c_k&=a_k-A_{k-1}d_k\\&=a_k-A_{k-1}A_{k-1}^+a_k\\&=(I-P_{\mathscr{R}(A_{k-1}),\mathscr{N}(A_{k-1}^H)})a_k\\&=P_{\mathscr{N}(A_{k-1}^H),\mathscr{R}(A_{k-1})}a_k\\&\qquad k=2,\cdots,n\end{aligned}$$

定理3.5.8（Grevijje） 设 $A\in \mathbf{C}^{m\times n}$，则$A_k(2\leqslant k\leqslant n)$的Moors-Penrose广义逆为

$$A_k^+=[A_{k-1}a_k]=\begin{bmatrix}A_{k-1}^+-d_kb_k^H\\b_k^H\end{bmatrix}$$

其中

$$b_k^H=\begin{cases}c_k^+, & c_k\neq 0\\(1+d_k^Hd_k)^{-1}d_k^HA_{k-1}^+, & c_k=0\end{cases}$$

注意，若$c_k=0$且$d_k=0$，则由d_k及c_k的定义可知$a_k=0$，并且$b_k=0$，这显然是预期的结果。

证明 现将$A_k^+=[A_{k-1}a_k]^+$假设为下面的分块形式

$$A_k^+=\begin{bmatrix}B_k\\b_k^H\end{bmatrix}$$

其中b_k^H为A_k^+的第k行。将A_k与A_k^+相乘可得

$$A_k A_k^+ = A_{k-1} B_k + a_k b_k^H$$

由A_k的分块表示及引理3.3.10，便知

$$\mathscr{N}(A_{k-1}^+) = \mathscr{N}(A_{k-1}^H) \supset \mathscr{N}(A_k^H) = \mathscr{N}(A_k^+)$$
$$= \mathscr{N}(A_k A_k^+)$$

由于$AP_{L,M} = A$当且仅当$\mathscr{N}(A) \supset M$，故有

$$A_{k-1}^+ A_k A_k^+ = A_{k-1}^H$$

此外，再由引理3.3.10可知

$$\mathscr{R}(A_k^+) = \mathscr{R}(A_k^H)$$

由A_k及A_k^+的分块表示，我们可推得

$$\mathscr{R}(B_k) \subset \mathscr{R}(A_{k-1}^H) = \mathscr{R}(A_{k-1}^+) = \mathscr{R}(A_{k-1}^+ A_{k-1})$$

又由于$P_{L,M}A = A$当且仅当$\mathscr{R}(A) \subset L$，故有

$$A_{k-1}^+ A_{k-1} B_k = B_k$$

把式$A_k A_k^+ = A_{k-1} B_k + a_k b_k^H$的两端左乘以$A_{k-1}^+$便得

$$A_{k-1}^+ = B_k + A_{k-1}^+ a_k b_k^H$$
$$= B_k + d_k b_k^H$$

故有

$$[A_{k-1} a_k]^+ = \begin{bmatrix} A_{k-1}^+ - d_k b_k^H \\ b_k^H \end{bmatrix}$$

该式中的b_k^H是待定向量。下面就d_k是否落在$\mathscr{R}(A_{k-1})$中，也就是视c_k是否为零来加以讨论。

若$c_k \neq 0$，利用$B_k = A_{k-1}^+ - d_k b_k^H$，便可得到

$$A_k A_k^+ = A_{k-1} B_k + a_k b_k^H$$
$$= A_{k-1} A_{k-1}^+ + (a_k - A_{k-1} d_k) b_k^H$$
$$= A_{k-1} A_{k-1}^+ + c_k b_k^H$$

由于$A_k A_k^+$具有 Hermite性，故$c_k b_k^H$也具有 Hermite性，因此应有

$$b_k^H = \delta c_k^H$$

式中的δ为某一实数。再根据A_k的分块表示及c_k的定义可得

$$A_k = A_k A_k^+ A_k = [A_{k-1} + c_k b_k^H \quad A_{k-1} a_k - c_k$$
$$+ (b_k^H a_k) c_k]$$

将该式再与 $A_k=[A_{k-1}a_k]$ 比较，特别是最后一列，便得到 $b_k^H a_k=1$，因为$c_k\neq 0$。并且还有

$$1=b_k^H a_k=\delta c_k^H a_k=\delta a_k^H P_{\mathscr{N}(A_{k-1}^H),\mathscr{R}(A_{k-1})}a_k$$
$$=\delta a_k^H P^2_{\mathscr{N}(A_{k-1}^H),\ \mathscr{R}(A_{k-1})}a_k=\delta c_k^H c_k$$

故有

$$b_k^H=\delta c_k^H=c_k^H/c_k^H c_k=c_k^+$$

若$c_k=0$，据 c_k 的定义知 $a_k\in\mathscr{R}(A_{k-1})$，因而 $\mathscr{R}(A_k)=\mathscr{R}(A_{k-1})$。由$A_k^+$的行分块及引理3.3.10可知

$$\mathscr{N}(b_k^H)\supset\mathscr{N}(A_k^+)=\mathscr{N}(A_k^H)=\mathscr{N}(A_{k-1}^H)=\mathscr{N}(A_{k-1}^+)$$
$$=\mathscr{N}(A_{k-1}A_{k-1}^+)$$

因此有

$$b_k^H A_{k-1}A_{k-1}^+=b_k^H$$

再根据

$$A_k=[A_{k-1}a_k],\ [A_{k-1}a_k]^+=\begin{bmatrix} A_{k-1}^+-d_k b_k^H \\ b_k^H \end{bmatrix}$$

可得

$$A_k^+A_k=\begin{bmatrix} A_{k-1}^+A_{k-1}-d_k b_k^H A_{k-1} & (1-\alpha)d_k \\ b_k^H A_{k-1} & \alpha \end{bmatrix}$$

式中$\alpha=b_k^H a_k$为某一实标量，因为它是Hermite矩阵$A_k^+A_k$的对角元,同时据$A_k^+A_k$的Hermite性还有$b_k^H A_{k-1}=(1-\alpha)d_k^H$。从而得到

$$b_k^H=b_k^H A_{k-1}A_{k-1}^+=(1-\alpha)d_k^H A_{k-1}^+$$
$$\alpha=b_k^H a_k=(1-\alpha)d_k^H A_{k-1}^+a_k=(1-\alpha)d_k^H d_k$$

上式两端同时加上$(1-\alpha)$，便有

$$(1-\alpha)(1+d_k^H d_k)=1$$
$$1-\alpha=(1+d_k^H d_k)^{-1}$$

这样便最后证明了

$$b_k^H=(1-\alpha)d_k^H A_{k-1}^+=(1+d_k^H d_k)^{-1}d_k^H A_{k-1}^+$$

在实用时，多数情况不是计算A^+本身，而是结合某一 $y\in\mathbf{C}^m$

计算A^+y。根据Greville的方法可以这样解决：令

$$\tilde{A}=[A \quad y]$$

则据Greville关于A_k^+的递推公式有

$$A_k^+\tilde{A}=\begin{bmatrix} A_{k-1}^+\tilde{A}-d_k b_k^H\tilde{A} \\ b_k^H\tilde{A} \end{bmatrix}$$

另外再比较$d_k=A_{k-1}^+a_k$与 $A_{k-1}^+\tilde{A}=[A_{k-1}^+A\ A_{k-1}^+y]$ 知 d_k 是$A_{k-1}^+\tilde{A}$的第k列，$k=2，3，\cdots n$。因此，为得到 $A_{k-1}^+\tilde{A}$ ，只要求得$b_k^H\tilde{A}$就够了。很显然，当$c_k=0$时，由Greville的公式知

$$b_k^H\tilde{A}=(1+d_k^H d_k)^{-1}d_k^+A_{k-1}^+\tilde{A}$$

而当$c_k\neq 0$时，有

$$b_k^H\tilde{A}=(c_k^H c_k)^{-1}c_k^H\tilde{A}$$

式中$c_k^H\tilde{A}$的第s列（$s=1，2，\cdots n$)就是 $c_k^H a_s$ ，而在定义的$c_k=a_k-A_{k-1}A_{k-1}^+a_k$两端左乘$c_k^H$，便得到

$$\begin{aligned} c_k^H c_k &= c_k^H a_k - a_k^H A_{k-1}A_{k-1}^+a_k \\ &= c_k^H a_k-(a_k^H-a_k^H A_{k-1}A_{k-1}^+)A_{k-1}A_{k-1}^+a_k \\ &= c_k^H a_k \end{aligned}$$

这表明，$b_k^H\tilde{A}$的第s个分量就是$c_k^H a_s/c_k^H a_k$， $s=1，2，\cdots n$。

用Greville方法计算矩阵的广义逆，在每次递推时加进一列。这种方法在某些情况下是方便的。例如在最小二乘意义下，用多项式

$$\sum_{j=0}^{k}x_j t^j$$

来逼近给定的实数$y(t)$，也就是确定$\{x_j\}$，使得在给定的m个点$(t_1，y_1)$，$(t_2，y_2)$，…，$(t_m，y_m)$上误差

$$\left(\sum_{i=1}^{m}\left|\sum_{j=0}^{k}x_j t_i^j-y_i\right|^2\right)^{1/2}\leqslant\varepsilon$$

若记$y=(y_1，y_2，\cdots，y_m)^T$，$a_{ij}=t_i^{j-1}$，且

$$A_{k+1}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\,k+1} \\ a_{21} & a_{22} & \cdots & a_{2\,k+1} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{m\,k+1} \end{pmatrix}=[a_1 a_2 \cdots a_{k+1}]$$

则上述问题归结为确定$k(\varepsilon)$及$x=(x_0, \cdots, x_k)\in \mathbf{R}^{k+1}$使得

$$\| A_{k+1}x-y\| \leqslant \varepsilon$$

根据$M-P$广义逆的性质就可知道$x=A_{k+1}^{+}y$使欧氏范数$\| A_{k+1}x-y\|$达到极小。如果在给定ε的情况下仍有

$$\| A_{k+1}A_{k+1}^{+}y-y\| > \varepsilon$$

则只有通过提高多项式的次数k才能解决。这时对$A_{k+2}=[A_{k+1}, a_{k+2}]$来重复上述的计算过程,其中$a_{k+2}=(t_1^{k+1}, t_2^{k+1}, \cdots, t_m^{k+1})^T$。如果还不行，则继续下去，直到满足为止。这就体现了Greville方法的特点。不过，在某些情况下，上述的Greville方法不好直接使用，所以有人又研究了一种对偶的形式。例如在参数的最小二乘估计中，常会遇到求解下述方程组的问题：

$$\sum_{j=1}^{n} a_{ij}x_j=y_i, \quad i=1, 2, \cdots, k-1$$

其解当然是$x=A_{k-1}^{+}y$。式中的

$$x=(x_1, x_2, \cdots, x_n)^T, \ y=(y_1, y_2, \cdots, y_{k-1})^T$$

$$A_{k-1}=(a_{ij})$$

如果观测数据由原来的 $k-1$ 个增加到 k 个时，相应的矩阵A由$A_{k-1}\in \mathbf{R}^{(k-1)\times n}$ 变为$A_k\in \mathbf{R}^{k\times n}$，即矩阵增加了一行而不是一列。这时为求$A_k^{+}$就必须采用与上述方法对偶的形式。为此，令

$$A_k=\begin{bmatrix} A_{k-1} \\ a_k^H \end{bmatrix}, \qquad a_k^H \in \mathbf{C}^{1\times n}$$

类似的定义

$$d_k^H=a_k^H A_{k-1}^{+}$$

$$c_k^H=a_k^H-d_k^H A_{k-1}$$

则对上面的A_k的分块表示取共轭转置得

$$A_k^H=[A_{k-1}^H \ a_k]$$

再应用定理3.5.8便可得到

$$A_k^+=[A_{k-1}^+-b_kd_k^H\ b_k]$$

其中

$$b_k=\begin{cases}(c_k^H)^+, & c_k\neq 0\\ (1+d_k^H d_k)^{-1}A_{k-1}^+d_k, & c_k=0\end{cases}$$

这便是定理3.5.8的对偶形式。特别是对给定的 $y_{(k)}\in\mathbf{C}^k$ 需要计算 $\hat{x}_{(k)}=A_k^+y_{(k)}$，而不是 A_k^+ 本身，则类似地有：对 $y_{(k)}$ 分块

$$y_{(k)}=\begin{bmatrix} y_{(k-1)}\\ y_k\end{bmatrix}$$

且令

$$\hat{x}_{(k)}=A_k^+y_{(k)},\quad \hat{x}_{(k-1)}=A_{k-1}^+y_{(k-1)}$$

则有

$$\hat{x}_{(k)}=\hat{x}_{(k-1)}+(y_k-a_k^H\ \hat{x}_{(k-1)})b_k$$

其中 b_k 由前面的式子给出。

第六节　广义逆的应用举例

本节给出一些广义逆的应用例子，以说明它的有用性。

例　在求解非线性方程组中的一个应用。

这里先介绍 {2}-逆的一个应用，因为在前面很少提到 {2}-逆的作用。设非线性方程组为

$$f_i(x_1, x_2, \cdots, x_n)=0,\ i=1, 2, \cdots, m$$

若记 $x=(x_1, x_2, \cdots, x_n)^T\in\mathbf{R}^n$，$F=(f_1, f_2, \cdots, f_m)^T$，则上述方程组可简记为 $F(x)=0$。解非线性方程组 $F(x)=0$ 的一个常用方法是Newton法。当 $m=n$ 时有

$$x_{(k+1)}=x_{(k)}-[F'(x_{(k)})]^{-1}F(x_{(k)}),\quad k=0, 1, \cdots$$

其中 $x_{(k)}$ 表示 $x=(x_1, x_2, \cdots, x_n)^T$ 的第 k 次迭代值而 $x_{(0)}$ 表示初始值，$F'(x_{(k)})$ 表示 $F(x)$ 在 $x_{(k)}$ 的一阶导数，它是一个Jacobi矩阵

$$F'(x_{(k)}) = \left[\frac{\partial f_i}{\partial x_j}\right]_{x=x_{(k)}}$$

有时为简单起见，乾脆取$F'(x_{(k)})=E'(x_{(0)})$。

如果$F'(x_{(k)})$不能保证对每个$k=0，1，\cdots$都是非奇异的，或者方程的个数m不等于未知数的个数n，即 $F'(x_{(k)})$ 不是方阵，上面的Newton迭代程序就不能使用。这时自然会想到用矩阵$F'(x_{(k)})$的广义逆来处理，即构造一个收敛到解 x 的迭代序列$\{x_{(k)}\}$。这里采用的是与向量相容的矩阵范数(见第四章第一节）$\|\cdot\|$。并定义空间$\mathbf{C}^n$中的以$x_{(0)}$为球心，以r为半径的开和闭球如下：

$$B(x_{(0)};\ r)=\{x\in\mathbf{C}^n \mid \|x-x_{(0)}\|<r\}$$

$$\overline{B}(x_{(0)};\ r)=\{x\in\mathbf{C}^n \mid \|x-x_{(0)}\|\leqslant r\}$$

则有下述结论。

定理3.6.1 设 $F: B(x_{(0)};\ r)\longrightarrow \mathbf{C}^m$ 是一个映射，其中$x_{(0)}\in\mathbf{C}^n$，$r>0$；存在$A\in\mathbf{C}^{m\times n}$，$T\in\mathbf{C}^{n\times m}$，$\varepsilon>0$，$\delta>0$，使得对任意的$u$，$v\in B(x_{(0)};\ r)$都满足

1. $\|F(u)-F(v)-A(u-v)\|\leqslant\varepsilon\|u-v\|$，
2. $TAT=T$，
3. $\varepsilon\|T\|=\delta<1$，
4. $\|T\|\ \|F(x_{(0)})\|<(1-\delta)r$，

则迭代序列

$$x_{(k+1)}=x_{(k)}-TF(x_{(k)})$$

一定收敛于满足

$$TF(x)=0$$

的点

$$x_*\in\overline{B}(x_{(0)};\ r)$$

证明 首先据条件1可推出映射F是连续的，因为有

$$\|F(u)-F(v)\|\leqslant\|A\|\ \|u-v\|+\varepsilon\|u-v\|$$

如果我们能够证明不等式

$$\| x_{(k+1)} - x_{(k)} \| \leqslant \delta^k (1-\delta) r, \quad k=0, 1, 2, \cdots$$

则立即据$\delta<1$知$\{ x_{(k)} \}$是一个Cauchy序列，由于 $\mathbf{C}^n$ 是完备的空间，从而推知它是收敛的，不妨设$\lim\limits_{k\to\infty} x_{(k)} = x_* \in \mathbf{C}^n$。而据$TF(x)$的连续性（在开球$B(x_{(0)}; r)$上），则有

$$TF(x_*) = TF(\lim_{k\to\infty} x_{(k)}) = \lim_{k\to\infty} TF(x_{(k)})$$

$$= \lim_{k\to\infty} (x_{(k+1)} - x_{(k)}) = 0$$

据假定证明了的不等式还可推出

$$\| x_{(k)} - x_{(0)} \| = \| x_{(k)} - x_{(k-1)} + x_{(k-1)} - \cdots$$

$$+ x_{(1)} - x_{(0)} \| \leqslant \| x_{(k)} - x_{(k-1)} \|$$

$$+ \| x_{(k-1)} - x_{(k-2)} \| + \cdots + \| x_{(k)} - x_{(0)} \|$$

$$\leqslant \sum_{i=0}^{k-1} (1-\delta) \delta^i r = (1-\delta^k) r$$

这表明$x_{(k)} \in B(x_{(0)}; r)$，$k=0, 1, 2, \cdots$，对上面不等式关于$k \longrightarrow \infty$求极限，便有

$$\| x_* - x_{(0)} \| \leqslant r$$

即证明了$x_* \in \bar{B}(x_{(0)}; r)$。到此证明了定理中的全部结论。留下来的是要证明不等式对一切k 成立。现在用归纳法证之。首先根据条件 4 有

$$\| x_{(1)} - x_{(0)} \| = \| -TF(x_{(0)}) \| \leqslant \| T \| \| F(x_{(0)}) \|$$

$$< (1-\delta) r$$

这表明$k=0$时不等式成立。现假定对$k-1$也成立，即有

$$\| x_{(k)} - x_{(k-1)} \| \leqslant \delta^{k-1} (1-\delta) r$$

我们来证对k 也是成立的：

$$\| x_{(k+1)} - x_{(k)} \| \leqslant \delta^k (1-\delta) r$$

因为

$$x_{(k+1)} - x_{(k)} = -TF(x_{(k)})$$

$$= -TF(x_{(k-1)}) - T[F(x_{(k)}) - F(x_{(k-1)})]$$

$$= T[A(x_{(k)} - x_{(k-1)}) - F(x_{(k)}) + F(x_{(k-1)})]$$

上面最后一个等式用了迭代关系式及条件 2 。对上面等式两端再取范数并利用条件1及3和归纳假设，便有所欲要的不等式：

$$\begin{aligned}\| x_{(k+1)} - x_{(k)} \| &\leqslant \| T \| \ \| F(x_{(k)}) - F(x_{(k-1)}) \\ &\quad - A(x_{(k)} - x_{(k-1)}) \| \\ &\leqslant \| T \| \varepsilon \| x_{(k)} - x_{(k-1)} \| \\ &= \delta \ \| x_{(k)} - x_{(k-1)} \| \\ &\leqslant \delta^k (1-\delta) r\end{aligned}$$

关于这个定理有几点说明：

1. 如果我们考虑的映射F，有$A \in \mathbf{C}^{m \times n}$存在使

$$\lim_{u \to v} \frac{\| F(u) - F(v) - A(u-v) \|}{\| u - v \|} = 0$$

则称F在v是Fréchet可微的，且称A是F在v的Fréchet导数。可以看出，前面定理中的A是一个近似的导数，因为并没有要求定理中的F是Fréchet可微的。

2. 迭代序列$\{ x_{(k)} \}$的极限x_*只是说为方程$TF(x)=0$的一个解，并没有断言它是 $F(x)=0$ 的解。在T是列满秩的情况下，$TF(x)=0$与$F(x)=0$等价，这时x_*便是$F(x)=0$的解。由于T是A的$\{2\}$-逆，它的秩介于零和$\mathrm{rank}(A)$之间，所以x_*究竟在多大的程度上可作为$F(x)=0$的解，取决于$\{2\}$-逆T的秩和其它的特征。最坏的选择是$T=0$，这时每一个初值$x_{(0)}$都是方程$TF(x)=0$的解，并且迭代过程在$x_{(0)}$处停止。

3. 对于$T \neq 0$，条件 4 表明$x_{(0)}$作为$Fx=0$的近似解误差界为

$$\| F(x_{(0)}) \| \leqslant \frac{(1-\delta)r}{\| T \|}$$

4. 值得注意的是条件1只要对满足$u - v \in \mathscr{R}(T)$的u，v成立就够了，且迭代序列$\{ x_{(k)} \}$的极限

$$x_* \in \overline{B}(x_{(0)};\ r) \cap \{ x_0 + \mathscr{R}(T) \}$$

由此可以有

推论3.6.2 设$F:B(x_{(0)};\ r)\longrightarrow\mathbf{C}^m$是一个映射，其中$x_{(0)}\in\mathbf{C}^n$，$r>0$，$\varepsilon>0$，$\delta>0$，$\eta>0$。若对任意的$x\in\overline{B}(x_{(0)};r)$矩阵

$$A_x\in\mathbf{C}^{m\times n},\qquad T_x\in\mathbf{C}^{n\times m}$$

对所有的u，$v\in B(x_{(0)};\ r)$都满足

$$\|F(u)-F(v)-A_v(u-v)\|\leqslant\varepsilon\|u-v\|$$
$$T_uA_uT_u=T_u$$
$$\|(T_u-T_v)F(v)\|\leqslant\eta\|u-v\|$$
$$\varepsilon\|T_u\|+\eta\leqslant\delta<1$$
$$\|T_{x_{(0)}}\|\,\|F(x_{(0)})\|<(1-\delta)r$$

则迭代序列

$$x_{(k+1)}=x_{(k)}-T_{x_{(k)}}F(x_{(k)})\qquad k=0,\ 1,\ \cdots$$

一定收敛于满足

$$T_{x_*}F(x)=0$$

的

$$x_*\in\overline{B}(x_{(0)};\ r)$$

证明 与定理3.6.1类似，略。

例 在线性规划中的一个应用。

现在介绍{1}-逆在区间线性规划中的一个应用。我们把线性规划问题

$$\max\{c^Tx\,|\,a\leqslant Ax\leqslant b\}\qquad(*)$$

叫做区间线性规划，或双边约束线性规划。其中$a,b\in\mathbf{R}^m$，$c\in\mathbf{R}^n$，$A\in\mathbf{R}^{m\times n}$。任何一个具有有界约束的线性规划都可以写成这种形式。下边仅就$A\in R_m^{m\times n}$的特殊情形讨论，并用A的{1}-逆给出最优解。当然，对一般的情况也可给出结论。为此，当集合

$$F=\{x\in\mathbf{R}^n\,|\,a\leqslant Ax\leqslant b\}\neq\phi$$

时，称问题(*)是相容的，并把F中的每一个元素都叫做问题(*)的可行解。如果还满足

$$\max\{c^T x \mid x\in F\}<\infty$$

则又称问题(*)是有界的。这时问题(*)的最优解就是满足

$$c^T x_0=\max\{c^T x \mid x\in F\}$$

的可行解x_0。下面引理表明，问题(*)的有界性等价于$c\in\mathscr{R}(A^T)$

引理3.6.3 相容的（即$F\neq\phi$）区间规划(*)是有界的，当且仅当$c\in\mathscr{R}(A^T)$。

证明 如果我们在线性空间中把子集的相加定义为$A+B=\{a+b \mid a\in A,\ b\in B\}$，则问题(*)中的$F=F+\mathscr{N}(A)$，这是很显然的。因此有

$$\begin{aligned}\max\{c^T x \mid x\in F\}&=\max\{c^T x \mid x\in F+\mathscr{N}(A)\}\\&=\max\{(P_{\mathscr{R}(A^T),\ \mathscr{N}(A)}c\\&\quad+P_{\mathscr{N}(A),\ \mathscr{R}(A^T)}c)^T x \mid x\in F+\mathscr{N}(A)\}\\&=\max\{c^T P_{\mathscr{R}(A^T),\ \mathscr{N}(A)}x \mid x\in F\}\\&\quad+\max\{c^T x \mid x\in\mathscr{N}(A)\}\end{aligned}$$

而因为

$$\begin{aligned}&\max\{c^T P_{\mathscr{R}(A^T),\mathscr{N}(A)}x \mid x\in F\}\\&=\max\{c^T A^+Ax \mid a\leqslant Ax\leqslant b\}\end{aligned}$$

是有限的，所以，$\max\{c^T x \mid x\in F\}$是有限的当且仅当$\max\{c^T x \mid x\in\mathscr{N}(A)\}$是有限的。显然有

$$\max\{c^T x \mid x\in\mathscr{N}(A)\}\Longleftrightarrow c\in\mathscr{N}(A)^{\perp}$$

为了表述的方便，定义映射

$$\eta:\ \mathbf{R}^m\times\mathbf{R}^m\times\mathbf{R}^m\longrightarrow\mathbf{R}^m$$

$$\eta(u,\ v,\ w)\longmapsto\eta=(\eta_1,\ \eta_2,\ \cdots,\ \eta_m)^T$$

其中$u=(u_1,\ u_2,\ \cdots,\ u_m)^T$，$v=(v_1,\ v_2,\ \cdots,\ v_m)^T$，$w=(w_1\ w_2,\ \cdots,\ w_m)^T$。

$$\eta_i=\begin{cases}u_i, & w_i<0\\ v_i, & w_i>0\\ \lambda_i u_i+(1-\lambda_i)v_i, & w_i=0,\quad 0\leqslant\lambda_i\leqslant1\end{cases}$$

也就是说，η的第i个分量的取值取决于w_i的第i个分量的正负。若w的某个分量等于零时，则η的对应分量可以取以u，v对应分量为端点的区间上的任一点。下面的定理便给出了当A为行满秩时的问题(*)的全部最优解。

定理3.6.4 若问题(*)是相容且有界的，而A又是行满秩的，$A^{(1)}$表示A的任一{1}-逆，则(*)的最优解的通式为

$$x=A^{(1)}\eta(a,\ b,\ A^{(1)T}c)+y,\qquad \forall y\in\mathscr{N}(A)$$

证明 由于$A\in\mathbf{R}^{m\times n}$且$\mathrm{rank}(A)=m$，故$\mathscr{R}(A)=\mathbf{R}^m$。因此，任给$u\in\mathbf{R}^m$，都有$u=Ax$，从而满足该等式的$x$都可表示为

$$x=A^{(1)}u+y,\qquad y\in\mathscr{N}(A)$$

将该式代入$\max\{c^Tx\,|\,a\leqslant Ax\leqslant b\}$之中，并注意$c\perp y$，则有等价的表示

$$\begin{aligned}&\max\{c^TA^{(1)}u\,|\,a\leqslant u\leqslant b\}\\&=\max\{(A^{(1)T}c)^Tu\,|\,a\leqslant u\leqslant b\}\end{aligned}$$

由此很容易看出它的最优解为

$$u=\eta(a,\ b,\ A^{(1)T}c)$$

再代回x的表达式，便得到

$$x=A^{(1)}\eta(a,\ b,\ A^{(1)T}c)-y,\quad \forall y\in\mathscr{N}(A)$$

例 在控制理论的极点配置方面的应用：在多变量控制论理中，有一个重要的问题是利用线性状态反馈或输出反馈来配置闭环系统的极点，使之具有预先指定的特征值集合。除非是单输入系统，否则仅指定闭环的特征值并不能唯一地确定一个闭环系统。这种不唯一性使得状态/输出反馈在配置了特征值的同时，可以在某一个集合内自由地选取特征向量。

现在来看线性时不变多变量系统

$$\dot{x}=Ax+Bu$$
$$y=Cx$$

的输出反馈问题，其中$A\in\mathbf{R}^{n\times n}$，$B\in\mathbf{R}_m^{n\times m}$，$C\in\mathbf{R}_l^{l\times n}$，$x=(x_1(t),x_2(t),\cdots,x_n(t))^T$，$\dot{x}=(\dot{x}_1(t),\ \cdots,\ \dot{x}_n(t))^T$。我们希望通过

输出反馈使得系统具有指定的闭环特征值与特征向量集合：

$$\{\lambda_1, \lambda_2, \cdots, \lambda_k\}, \quad \{s_1, s_2, \cdots, s_k\}$$

其中$\lambda_i \in \mathbf{C}$，$s_i \in \mathbf{C}^n$，$1 \leqslant k \leqslant n$。所谓特征值-特征向量可配置的概念，就是给定的系统如果存在输出反馈

$$u = Ky + v, \qquad K \in \mathbf{R}^{m \times l}$$

使得闭环系统

$$\dot{x} = (A + BKC)x + Bv$$

具有所指定的特征值与特征向量$\{\lambda_i\}$与$\{s_i\}$，这时就称之为(A,B,C)可配置的。现就λ_i为单实根的情形给出一个结论，而一般情形，类似的方法可以处理。

定理3.6.5 系统

$$\dot{x} = Ax + Bu, \qquad y = Cx$$

$$A \in \mathbf{R}^{n \times n}, \quad B \in \mathbf{R}_m^{n \times m}, \quad C \in \mathbf{R}_l^{l \times n}$$

关于指定的特征值-特征向量

$$\{\lambda_1, \lambda_2, \cdots, \lambda_k\}, \quad \{s_1, s_2, \cdots, s_k\}$$

是(A, B, C)可配置的，当且仅当

$$BB^{(1)}(S\Lambda - AS)(CS)^{(1)}CS = (S\Lambda - AS)$$

并且配置它们所用的反馈增益阵K的通式为

$$K = B^{(1)}(S\Lambda - AS)(CS)^{(1)} + Z[I - CS(CS)^{(1)}]$$

其中$B^{(1)}$及$(CS)^{(1)}$分别为B及CS的任一$\{1\}$-逆，这里$B^{(1)} = (B^TB)^{-1}B^T$，$Z \in \mathbf{C}^{m \times l}$是任意的，并且

$$S = [s_1, s_2, \cdots, s_k], \quad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_k)$$

证明 所谓闭环系统 $\dot{x} = (A + BKC)x + Bu$ 具有指定的特征值$\{\lambda_i\}$及特征向量$\{s_i\}$是指满足

$$(A + BKC)s_i = \lambda_i s_i, \qquad i = 1, 2, \cdots, k$$

又可改写为

$$BKCS = S\Lambda - AS$$

对于给定的 A，B，C，Λ，S，存在K满足上面方程的充要条件为（定理3.3.1）

$$BB^{(1)}(S\Lambda - AS)(CS)^{(1)}CS = S\Lambda - AS$$

在满足这一条件的前提下，K的通式为

$$K = B^{(1)}(S\Lambda - AS)(CS)^{(1)} + Z[I - CS(CS)^{(1)}]$$

这里用到了B是列满秩的条件，即用到$B^{(1)}B = I$。

由这个定理可以看出，一般来说，不是任意的系统(A, B, C)关于指定的 $\Lambda - S$ 都是可配置的；即使是可配置的，反馈增益阵K也不一定是唯一的，一般来说，K的集合构成矩阵空间 $\mathbf{C}^{m\times l}$ 中的一个仿射流形（子空间的一个平移）。这就使得我们有可能在某种意义下作出最优的选择，这本身也是一个很有意义的课题。

与控制理论密切相关的还有参数估计问题。统计估计理论也是广义逆矩阵应用得最好的一个领域，下面举一个简单的例子。

例 在参数估计中的一个应用：设参数估计问题中的量测方程为

$$y_i = h_i^T x + n_i, \qquad i = 1, 2, \cdots, k$$

其中y_i是第i个观测值，n_i为观测噪声，$x \in \mathbf{R}^n$是待估计的参数向量，$h_i \in \mathbf{R}^n$也是已知的结构参数向量。这里假定噪声具有下列统计特性：

$$E[n_i] = 0, \qquad E[n_i n_j] = q_i \delta_{ij}$$

其中 $E[\cdot]$ 表示均值算子（或取数学期望），δ_{ij} 为常用的 Kronecker 符号。限于篇幅，这里不加证明地给出用到的一个引理。

引理3.6.6 上述量测方程

$$y_i = y_i^T x + n_i, \qquad i = 1, 2, \cdots, k$$

的加权最小二乘最小范数解为

$$\hat{x} = A_k^+ Z_k$$

其中

$$A_k = \begin{bmatrix} q_1^{-\frac{1}{2}} h_1^T \\ \vdots \\ q_k^{-\frac{1}{2}} h_k^T \end{bmatrix} = \begin{bmatrix} \bar{h}_1 \\ \vdots \\ \bar{h}_k \end{bmatrix} \in \mathbf{R}^{k\times n},$$

$$Z_k=\begin{bmatrix} q_1^{-\frac{1}{2}}y_1 \\ \vdots \\ q_k^{-\frac{1}{2}}y_k \end{bmatrix} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_k \end{bmatrix} \in \mathbf{R}^{k\times 1}$$

若定义协方差阵为

$$P_k=E[(\hat{x}_k-A_k^+A_kx)(\hat{x}_k-A_k^+A_kx)^T]$$

则还有

引理3.6.7 对于上面所定义的P_k，则有

$$P_k=A_k^+(A_k^+)^T=(A_k^TA_k)^+$$

证明 令$Q_k=\mathrm{diag}(q_1, q_2, \cdots, q_k)$,

$$Y_k=\begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}, \quad N_k=\begin{bmatrix} n_1 \\ \vdots \\ n_k \end{bmatrix}, \quad H_k=\begin{bmatrix} h_1^T \\ \vdots \\ h_k^T \end{bmatrix}$$

则有 $$A_k=Q_k^{-\frac{1}{2}}H_k, \quad Z_k=Q_k^{-\frac{1}{2}}Y_k,$$

$$\begin{aligned}\hat{x}_k-A_k^+A_kx&=A_k^+Z_k-A_k^+A_kx \\ &=A_k^+Q_k^{-\frac{1}{2}}(Y_kH_kx) \\ &=A_k^+Q_k^{-\frac{1}{2}}N_k\end{aligned}$$

$$\begin{aligned}P_k&=A_k^+Q_k^{-\frac{1}{2}}E[N_kN_k^T]Q_k^{-\frac{1}{2}}(A_k^+)^T \\ &=A_k^+Q_k^{-\frac{1}{2}}Q_kQ_k^{-\frac{1}{2}}(A_k^+)^T=A_k(A_k^+)^T\end{aligned}$$

根据定理 3.2.5 有 $A_k^+=(A_k^TA_k)^+A_k^T$，两端取转置又有$(A_k^+)^T=A_k[(A_k^TA_k)^+]^T=A_k(A_k^TA_k)^+$，从而有

$$A_k^+(A_k^+)^T=(A_k^TA_k)^+A_k^TA_k(A_k^TA_k)^+=(A_k^TA_k)^+$$

这样就完成了引理3.6.7的证明。

定理3.6.8 任给矩阵U, $V\in\mathbf{R}^{m\times n}$，总有

$$\begin{aligned}(UU^T+VV^T)^+=&[I-(C^+)^TV^T](U^+)^T \\ &\times[I-U^+V(I-C^+C)KV^TU^+] \\ &\times U^+(I-VC^+)+(C^+)^TC^+\end{aligned}$$

式中$C=(I-UU^+)V$,

$$K=[I+(I-C^+C)V^TU^+(U^+)^TUV(I-C^+C)]^{-1}$$

这个定理的证明偏长，略。有兴趣的读者可参阅R.E.Chine,

Representation for generalized inverse of sums of matrices.J.SIAM,V12.pp97-114，1965。

由上面定理可得到如下重要的推论：

推论3.6.9 在新增加一个量测值y_{k+1}后，协方差阵

$$P_{k+1}=\left[(A_k^T \vdots \tilde{h}_{k+1})\begin{pmatrix} A_k \\ \tilde{h}_{k+1}^T \end{pmatrix}\right]^+$$

可以写为

$$P_{k+1}=[I_n-(D_k^+)^T\tilde{h}_{k+1}][P_k-\tilde{h}_{k+1}(1-D_k^+D_k)K\tilde{h}_{k+1}P_k]$$
$$\times(I_n-\tilde{h}_{k+1}D_k^+)+(D_k^+)^TD_k^+$$

其中

$$D_k=(I_n-A_k^+A_k)\tilde{h}_{k+1}=(I_n-P_kA_k^TA_k)\tilde{h}_{k+1}$$
$$K=[1-(1-D_k^+D_k)\tilde{h}_{k+1}^TP_k\tilde{h}_{k+1}(1-D_k^+D_k)]^{-1}$$

有了上面的准备，就可以考虑参数的递推估计问题。兹分下列三种情况讨论：

1. 所有参数都是可观测的，且$\mathrm{rank}(A_k)=n$；
2. 至少有一个参数是不可观测的，且$\mathrm{rank}(A_k)=r<n$且$\mathrm{rank}(A_{k+1})=\mathrm{rank}(A_k)$；
3. 至少有一个参数是不可观测的，但有$\mathrm{rank}(A_{k+1})>\mathrm{rank}(A_k)$。

引理3.6.10 设

$$A_{k+1}=\begin{pmatrix} A_k \\ \tilde{h}_{k+1}^T \end{pmatrix}, \quad A_k\in\mathbf{R}^{m\times n}, \quad \tilde{h}_{k+1}^T\in\mathbf{R}^{1\times n} \quad m>n$$

则当且仅当$\mathrm{rank}(A_{k+1})=\mathrm{rank}(A_k)$时有

$$D_k=(I_n-A_k^+A_k)\tilde{h}_{k+1}=0$$

证明 当$\mathrm{rank}(A_{k+1})=\mathrm{rank}(A_k)$时，必存在$\alpha\in\mathbf{R}^k$使得

$$\tilde{h}_{k+1}^T=\alpha^TA_k$$

于是便有

$$D_k=(I_n-A_k^+A_k)\tilde{h}_{k+1}=[I_n-(A_k^+A_k)^T]A_k^T\alpha$$
$$=[A_k(I_n-A_k^+A_k)]^T\alpha=0$$

反之，若$D_k=0$，则有

$$\bar{h}_{k+1}=A_k^+A_k\bar{h}_{k+1}$$

这表明$\bar{h}_{k+1}\in\mathscr{R}(A_k^+A_k)$，又据引理3.3.10中2，即

$$\mathscr{R}(A^+A)=\mathscr{R}(A^+)=\mathscr{R}(A^H)$$

可得到

$$\bar{h}_{k+1}\in\mathscr{R}(A_k^T)$$

从而由$A_{k+1}^T=(A_k^T\bar{h}_{k+1})$可知 $\mathrm{rank}(A_{k+1}^T)=\mathrm{rank}(A_k^T)$ ，当然也有$\mathrm{rank}(A_{k+1})=\mathrm{rank}(A_k)$。

定理3.6.11 对于上述的三种情况，我们有如下的递推估计式

$$\hat{x}_{k+1}=\hat{x}_k+P_{k+1}\bar{h}_{k+1}(\tilde{y}_{k+1}-\bar{h}_{k+1}^T\hat{x}_k)$$

其中对情形1，2有

$$P_{k+1}=P_k-P_k\bar{h}_{k+1}K\bar{h}_{k+1}^TP_k$$

而对情形 3 则有

$$P_{k+1}=[I_n-(D_k^+)^T\bar{h}_{k+1}^T]P_k[I_n-\bar{h}_{k+1}D_k^+]+(D_k^+)^TD_k^+$$

式中

$$K=(1+\bar{h}_{k+1}^TP_k\bar{h}_{k+1})^{-1}$$

$$D_k=(I_n-A_k^+A_k)\bar{h}_{k+1}=(I_n-P_kA_k^TA_k)\bar{h}_{k+1}$$

证明 情形1 由于此时A_k是列满秩的，故有

$$A_k^+A_k=I_n$$

从而有$D_k=0$。于是推论3.6.9的P_{k+1}及K分别简化为

$$P_{k+1}=P_k-P_k\bar{h}_{k+1}K\bar{h}_{k+1}^TP_k$$

$$K=(1+\bar{h}_{k+1}^TP_k\bar{h}_{k+1})^{-1}$$

又由于

$$\begin{aligned}P_{k+1}\bar{h}_{k+1}&=P_k\bar{h}_{k+1}-P_k\bar{h}_{k+1}K\bar{h}_{k+1}^TP_k\bar{h}_{k+1}\\&=P_k\bar{h}_{k+1}\left(1-\frac{\bar{h}_{k+1}^TP_k\bar{h}_{k+1}}{1+\bar{h}_{k+1}^TP_k\bar{h}_{k+1}}\right)\\&=P_k\bar{h}_{k+1}K\end{aligned}$$

另据引理3.6.7有

$$A_k^+=(A_k^TA_k)^+A_k^T,\quad P_k=(A_k^TA_k)^+$$

$$\hat{x}_k=A_k^+Y_k=(A_k^TA_k)^+A_k^TY_k$$
$$=P_kA_k^TY_k$$

从而有

$$\hat{x}_{k+1}=P_{k+1}A_{k+1}^TY_{k+1}=P_{k+1}[A_k^T\tilde{h}_{k+1}]\begin{bmatrix}Y_k\\ \tilde{y}_{k+1}\end{bmatrix}$$
$$=P_{k+1}[A_k^TY_k+\tilde{h}_{k+1}\tilde{y}_{k+1}]$$
$$=[P_k-P_k\tilde{h}_{k+1}K\tilde{h}_{k+1}^TP_k][A_k^TY_k+\tilde{h}_{k+1}\tilde{y}_{k+1}]$$
$$=\hat{x}_k+P_k\tilde{h}_{k+1}\tilde{y}_{k+1}-P_k\tilde{h}_{k+1}K\tilde{h}_{k+1}^T\hat{x}_k$$
$$-P_k\tilde{h}_{k+1}K\tilde{h}_{k+1}^TP_k\tilde{h}_{k+1}y_{k+1}$$
$$=\hat{x}_k+P_{k+1}\tilde{h}_{k+1}\tilde{y}_{k+1}-P_{k+1}\tilde{h}_{k+1}\tilde{h}_{k+1}^T\hat{x}_k$$
$$=\hat{x}_k+P_{k+1}\tilde{h}_{k+1}(\tilde{y}_{k+1}-\tilde{h}_{k+1}^T\hat{x}_k)$$

情形 2　此时$\mathrm{rank}(A_{k+1})=\mathrm{rank}(A_k)$，据引理3.6.10有$D_k=0$，所以下面的证明便完全同情形1。

情形 3　此时$\mathrm{rank}(A_{k+1})>\mathrm{rank}(A_k)$，据引理3.6.10知

$$D_k=(I_n-A_k^+A_k)\tilde{h}_{k+1}\neq 0$$

但因为$D_k\in\mathbf{R}^{n\times 1}$，故$D_k^+=(D_k^TD_k)^{-1}D_k^T$，从而有

$$D_k^+D_k=1$$

则协方差阵的表达式为（推论3.6.9中的P_{k+1}）

$$P_{k+1}=[I_n-(D_k^+)^T\tilde{h}_{k+1}^T]P_k[I_n-\tilde{h}_{k+1}D_k^+]+(D_k^+)^TD_k^+$$

而参数估计

$$\hat{x}_{k+1}=P_{k+1}(A_k^TY_k+\tilde{h}_{k+1}\tilde{y}_{k+1})$$

可以进一步简化。先把 $D_k=(I_n-A_k^+A_k)\tilde{h}_{k+1}$ 代入$D_k^+=(D_k^TD_k)^{-1}D_k^T$中得到

$$D_k^+=[\tilde{h}_{k+1}^T(I_n-A_k^+A_k)^T(I_n-A_k^+A_k)\tilde{h}_{k+1}^T]^{-1}$$
$$\times\tilde{h}_{k+1}^+(I_n-A_k^+A_k)^T$$
$$=[\tilde{h}_{k+1}^T(I_n-A_k^+A_k)\tilde{h}_{k+1}]^{-1}\tilde{h}_{k+1}^T(I_n-A_k^+A_k)$$

上式两端右乘$\tilde{h}_{k+1}$便得到$D_k^+\tilde{h}_{k+1}=1$，从而有

$$P_{k+1}\tilde{h}_{k+1}=[I_n-(D_k^+)^T\tilde{h}_{k+1}^T]P_k[I_n-\tilde{h}_{k+1}D_k^+]\tilde{h}_{k+1}$$

$$+(D_k^+)^T D_k^+ \bar{h}_{k+1} = (D_k^+)^T$$

从上面推导出的D_k^+的表达式很容易看出有

$$D_k^+ A_k^T = 0$$

这样便又得到

$$\begin{aligned} P_{k+1} A_k^T &= [I_n - (D_k^+)^T \bar{h}_{k+1}^T] P_k A_k^T \\ &= P_k A_k^T - (D_k^+)^T \bar{h}_{k+1}^T P_k A_k^T \\ &= P_k A_k^T - P_{k+1} \bar{h}_{k+1} \bar{h}_{k+1}^T P_k A_k^T \end{aligned}$$

$$\begin{aligned} \hat{x}_{k+1} &= P_{k+1}(A_k^T Y_k + \bar{h}_{k+1} \bar{y}_{k+1}) \\ &= P_{k+1} A_k^T Y_k + P_{k+1} \bar{h}_{k+1} \bar{y}_{k+1} \\ &= P_k A_k^T Y_k - P_{k+1} \bar{h}_{k+1} \bar{h}_{k+1}^T P_k A_k^T Y_k \\ &\quad + P_{k+1} \bar{h}_{k+1} \bar{y}_{k+1} \\ &= \hat{x}_k + P_{k+1} \bar{h}_{k+1} (\bar{y}_{k+1} - \bar{h}_{k+1}^T \hat{x}_k) \end{aligned}$$

这样，就完全给出了利用广义逆进行估计的一个递推算法。

第七节　例题与习题

1．用不同于课文中的方法证明Moore-Penrose 广义逆的唯一性。

2．举一个明显的例子来说明集合$A\{2, 3, 4\}$非空。

3．证明引理3.1.1。

4．证明每一个方阵都有非奇异的$\{1\}$-逆。

5．设 $A=FHG$ ，其中F是列满秩， G 是行满秩，证明 $\text{rank}(A)=\text{rank}(H)$。

6．证明方阵A非奇异的充分必要条件是它有唯一的$\{1\}$-逆就是A^{-1}。

[证明]　任给 $x\in\mathscr{N}(A)$ ， $y\in\mathscr{N}(A^H)$ ， 分别加到某个$X\in A\{1\}$ 的任一列和任一行上，便可得 A 的另外的$\{1\}$-逆，从而说明$\{1\}$-逆的唯一性等价于

$$\mathscr{N}(A)=\{0\}, \mathscr{N}(A^H)=\{0\}$$

这表明方阵A是行满秩和列满秩，即A非奇异。

7. 证明$\mathscr{R}(AB)=\mathscr{R}(A)$的充分必要条件为

$$\operatorname{rank}(AB)=\operatorname{rank}(A)$$

$\mathscr{N}(AB)=\mathscr{N}(B)$的充分必要条件是

$$\operatorname{rank}(AB)=\operatorname{rank}(B)$$

[证明] 设$B\in\mathbf{C}^{m\times n}$。显然有$\mathscr{R}(AB)\subset\mathscr{R}(A)$，说明$\mathscr{R}(AB)$为$\mathscr{R}(A)$的子空间。据$\mathscr{R}(AB)=\mathscr{R}(A)$知

$$\dim\mathscr{R}(AB)=\dim\mathscr{R}(A)$$

而又知$\dim\mathscr{R}(AB)=\operatorname{rank}(AB)$，$\dim\mathscr{R}(A)=\operatorname{rank}(A)$，故必要性得证。反推回去，便证明充分性 。

很显然$\mathscr{N}(B)\subset\mathscr{N}(AB)$，即$\mathscr{N}(B)$为$\mathscr{N}(AB)$的子空间。据$\mathscr{N}(AB)=\mathscr{N}(B)$知，$\mathscr{N}(B)$与$\mathscr{N}(AB)$有相同的维数，而它们分别是方程

$$Bx=0 \qquad 与 \qquad ABx=0$$

的解空间，其维数分别是

$$n-\operatorname{rank}(B) \qquad 与 \qquad n-\operatorname{rank}(AB)$$

故命题是成立的。

8. 设$A\in\mathbf{C}^{m\times n}$的第$(i,\ j)$元素为1，其余元素皆为 0，试问$A\{1\}$是怎样的一类矩阵。

9. A同上题。试问A的Hermite 标准形是什么？证明把A化成Hermite标准形时的满秩方阵E，P 可取成置换阵，并且可取成最简单（即与同阶单位阵相比，对应元素的差异最小的）置换阵。在使用这种E和P之后，定理3.1.4给出的X 具有什么样的形式？它是否为$A\{1\}$的一般形式？

10. 上题的答案表明，对特别选定的E，P 只通过改变定理3.1.4中的L，是得不到A的所有的 $\{1\}$-逆的。因为在定理3.1.4中并没有要求 E，P 是置换阵。这里自然会提出这样的问题：可否通过取遍所有使

$$QAP=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \qquad (*)$$

的P和Q 来得到A 的所有{1}－逆？给定 $A\in \mathbf{C}_r^{m\times n}$ ，试证：$X\in A\{1\}$的充分必要条件是，对某L 及满足（＊）的非奇异阵P，Q有

$$X=P\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}Q$$

［证明］ 充分性已包含在定理 3.1.4 中 ，现看必要性。设$AXA=A$成立，据引理3.1.5，AX与XA都是秩为r的幂等阵，而幂等阵的特征值非0即1，且都与对角阵相似，不妨设

$$R^{-1}AXR=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times m},$$

$$P^{-1}XAP=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times n}$$

故有

$$\begin{aligned} R^{-1}AP &= R^{-1}AXAXAP \\ &=(R^{-1}AXR)(R^{-1}AP)(P^{-1}XAP) \\ &=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} R^{-1}AP\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

由此可以看出$R^{-1}AP$具有如下的形式：

$$R^{-1}AP=\begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix}, \qquad H\in \mathbf{C}_r^{r\times r}$$

说明H是非奇异的，故可选取

$$Q=\begin{bmatrix} H^{-1} & 0 \\ 0 & I_{m-r} \end{bmatrix}R^{-1}$$

使(＊)成立。而对于矩阵$P^{-1}XQ^{-1}$，容易看出：

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}(P^{-1}XQ^{-1})=QAPP^{-1}XQ^{-1}=QAXQ^{-1}$$

$$=\begin{bmatrix} H^{-1} & 0 \\ 0 & I_{m-r} \end{bmatrix}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} H & 0 \\ 0 & I_{m-r} \end{bmatrix}$$

$$=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

$$(P^{-1}XQ^{-1})\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}=P^{-1}XQ^{-1}QAP$$

$$=P^{-1}XAP=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

这表明

$$P^{-1}XQ^{-1}=\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}$$

对某一L成立，从而得到

$$X=P\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}Q$$

11. 设

$$A=\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

分别求出秩为1，2，3的{2}-逆。

12. 对习题8中的A，求其{1，2}-逆的一般形式，其Moore-Penrose逆A^+又是什么？

13. 对任意的$A\in \mathbf{C}^{m\times n}$，证明：

a) $(A^+)^+=A$

b) $(A^H)^+=(A^+)^H$

c) $(A^T)^+=(A^+)^T$

d) $A^+=(A^HA)^+A^H=A^H(AA^H)^+$

14. 设$a，b\in \mathbf{C}^n$，试证：

a) $a^+=(a^Ha)^+a^H$

b) $(ab^H)^+=(a^Ha)^+(b^Hb)^+ba^H$

15. 设H为幂等 Hermite 矩阵，证明$H^+=H$。

16. 试证：$H^+=H$的充分必要条件为，H^2是幂等Hermite矩阵且$\mathrm{rank}(H^2)=\mathrm{rank}(H)$。

17. 试证对角阵$D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$的 Moore-Penrose 逆为$D^+=\mathrm{diag}(d_1^+, d_2^+, \cdots, d_n^+)$。

18. 设U，V是酉阵且UAV有意义，则总有

$$(UAV)^+=V^HA^+U^H$$

19. 若$A=FG$为$A\in\mathbf{C}_r^{m\times n}$的满秩分解，即$F\in\mathbf{C}_r^{m\times r}$，$G\in\mathbf{C}_r^{r\times n}$，试证：

a) $G^{(i)}F^{(1)}\in A\{i\}$，$i=1, 2, 4$

b) $G^{(1)}F^{(j)}\in A\{j\}$，$j=1, 2, 3$

[提示] 利用等式$F^{(1)}F=GG^{(1)}=I_r$。

20. A，F，G如上题所设，试证：

$$A^+=G^+F^{(1,3)}=G^{(1,4)}F^+$$

21. 设A为一奇异方阵，$\{u_1, u_2, \cdots, u_k\}$与$\{v_1, v_2, \cdots, v_k\}$分别是$\mathscr{N}(A^H)$与$\mathscr{N}(A)$的标准正交基底，$\{\alpha_1, \alpha_2, \cdots, \alpha_k\}$为一组非零标量。则矩阵

$$A_0=A+\sum_{i=1}^{k}\alpha_iu_iv_i^H$$

是非奇异的，且其逆为

$$A_0^{-1}=A^++\sum_{i=1}^{k}\frac{1}{\alpha_i}v_iu_i^H$$

[提示] 利用$P_{\mathscr{N}(A^H),\mathscr{N}(A^H)^\perp}=\sum\limits_{i=1}^{k}u_iu_i^H=I_n-AA^+$

22. 设X_0是定理3.3.5中矩阵方程$AX=B$，$XD=E$的一个公共解，证明对任取的$A^{(1)}\in A\{1\}$，$D^{(1)}\in D\{1\}$，$Y\in\mathbf{C}^{m\times n}$，其公共解的通式为

$$X=X_0+(I-A^{(1)}A)Y(I-DD^{(1)})$$

［提示］ 先证明上式右端是公共解，然后设X为任一公共解，在上式中把Y取为$X-X_0$便得到该解。

23． 证明定理3.3.8及推论3.3.9。

24． 设矩阵

$$A=\begin{pmatrix} 0 & 2i & i & 0 & 4+2i & 1 \\ 0 & 0 & 0 & -3 & -6 & -3-3i \\ 0 & 2 & 1 & 1 & 4-4i & 1 \end{pmatrix}$$

证明$A\{1,3\}$的通式为

$$X=\begin{pmatrix} 0 & 0 & 0 \\ -10i & 3 & 9 \\ 0 & 0 & 0 \\ 2i & -12 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$+\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 & -1+2i & -\frac{1}{2}i \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & -1-i \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}Z$$

其中$Z\in \mathbf{C}^{6\times 3}$是任意的。

25． 对习题24中的矩阵A，证明$A\{1,4\}$的通式为

$$X=\frac{1}{276}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 20-18i & 42 \\ 0 & 10-9i & 21 \\ 0 & -29-9i & -9-27i \\ 0 & -2+4i & 24+30i \\ 0 & -29+30i & -36+3i \end{pmatrix}$$

$$+Z\begin{pmatrix} 1 & -\frac{1}{3}i & -i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

其中 $Z\in\mathbf{C}^{3\times 3}$是任意的。

26. 利用定理3.1.12及上两题的结果，计算A^{+}。

27. 利用定理3.3.6及定理3.3.8 给出定理 3.1.12 的另一证明。

[提示] 取$X=A^{+}$。

28. 设$A\in\mathbf{C}^{m\times n}$，$A^{(1,3,4)}\in A\{1,3,4\}$，利用习题22的结论证明

$$A\{1,3,4\}=\{A^{(1,3,4)}+(I-A^{(1,3,4)}A)Y(I-AA^{(1,3,4)})\mid Y\in\mathbf{C}^{n\times m}\}$$

29. 设$A\in\mathbf{C}^{m\times n}$，$A^{(1,2,3)}\in A\{1,2,3\}$，$A^{(1,2,4)}\in A\{1,2,4\}$，证明：

$$A\{1,2,3\}=\{A^{(1,2,3)}+(I-A^{(1,2,3)}A)ZA^{(1,2,3)}\mid Z\in\mathbf{C}^{n\times m}\}$$

$$A\{1,2,4\}=\{A^{(1,2,4)}+A^{(1,2,4)}Z(I-AA^{(1,2,4)})\mid Z\in\mathbf{C}^{m\times m}\}$$

30. 若E为幂等的 Hermite 矩阵。证明：$X\in E\{2,3,4\}$的充要条件是，X为幂等的 Hermite 矩阵且满足$\mathscr{R}(X)\subset\mathscr{R}(E)$。

[提示] 利用幂等阵的性质：

$$Ex=x\Longleftrightarrow x\in\mathscr{R}(E)$$

31. 设H为幂等的Hermite阵且是非负定的，其谱分解假定为

$$H=\sum_{i=1}^{k}\lambda_i P_i$$

证明：$X\in H\{2,3,4\}$的充要条件为

$$X=\sum_{i=1}^{k}\lambda_i^{+}G_i$$

其中 $G_i \in P_i\{2, 2, 4\}$。

［证明］ 充分性。由于谱算子P_i是幂等 Hermite 阵，据习题30知 $\mathscr{R}(G_i) \subset \mathscr{R}(P_i)$。因而据谱算子的性质有

$$P_i G_j = G_j P_i = 0, \quad i \neq j$$

再利用幂等阵的性质：

$$P_i x = x \Longleftrightarrow x \in \mathscr{R}(P_i)$$

还有

$$P_i G_j = G_j P_i = G_j, \quad i = 1, 2, \cdots, k$$

从而有

$$HX = \sum_{i=0}^{k}{}' G_i = XH$$

其中$\sum'$表示如果有 $\lambda_i = 0$，下标 i 便被省略的和式。据习题30知每个G_i皆为Hermite阵，因此$HX = XH$亦是 Hermite 阵。又因

$$G_i G_j = G_i P_j G_j = 0, \quad i \neq j$$

故有

$$XHX = \sum_{i=1}^{k} \lambda_i^{+} G_i^2 = \sum_{i=1}^{k} \lambda_i^{+} G_i = X$$

这表明$X \in H\{2, 3, 4\}$。

必要性。设$X \in H\{2, 3, 4\}$。利用H的谱算子 P_i 的性质有 $\sum_{i=1}^{k} P_i = I$，从而有

$$X = IXI = \sum_{i=1}^{k} \sum_{j=1}^{k} P_i X P_j \qquad (*)$$

$$HX = \sum_{i=1}^{k} \lambda_i P_i X = \sum_{i=1}^{k} \lambda_i X^H P_i$$

$$XH = \sum_{i=1}^{k} \lambda_i X P_i = \sum_{i=1}^{k} \lambda_i P_i X^H$$

其中用到$HX = XH$。对后两个等式分别左乘 P_s，右乘P_t，利用谱算子P_i的幂等性与正交性（或可分离性），便得到

$$\begin{matrix} \lambda_s P_s X P_t = \lambda_t P_s X^H P_t \\ \lambda_t P_s X P_t = \lambda_s P_s X^H P_t \end{matrix} \quad s, t = 1, 2, \cdots, k$$

将这两个等式分别相加和相减便得到

$$(\lambda_s + \lambda_t) P_s X P_t = (\lambda_s + \lambda_t) P_s X^H P_t$$

$$(\lambda_s - \lambda_t) P_s X P_t = -(\lambda_s - \lambda_t) P_s X^H P_t$$

注意到诸λ_i互异，且是非负的（因H是非负定的Hermite阵），因而当$s \neq t$时，$\lambda_s + \lambda_t$与$\lambda_s - \lambda_t$都不为零，于是有

$$P_s X P_t = P_s X^H P_t = -P_s X P_t = 0, \quad t \neq s$$

这样式（＊）便简化为

$$X = \sum_{i=1}^{k} P_i X P_i \qquad (**)$$

再利用H的谱分解式，便进一步得到

$$X = XHX = \sum_{i=1}^{k} \lambda_i X P_i X$$

利用式（＊＊）又可得到

$$P_s X P_s = \lambda_s P_s X P_s X P_s = \lambda_s (P_s X P_s)$$

据此可得：若$\lambda_s = 0$，则有$P_s X P_s = 0$。这时取

$$G_i = \lambda_i P_i X P_i, \quad i = 1, 2, \cdots, k$$

则式（＊＊）就成为

$$X = \sum_{i=1}^{k} \lambda_i^+ G_i, \quad (\lambda_i^+ = 0)$$

由于$\lambda_i = 0$，$G_i = 0$，显然$G_i \in P_i\{2, 3, 4\}$。而若$\lambda_i \neq 0$，则据前面的证明结果：（取$s = t = i$）

$$\lambda_i P_i X P_i = \lambda_i P_i X^H P_i$$

$$\lambda_i P_i X P_i = \lambda_i^2 (P_i X P_i)^2$$

即说明$G_i = \lambda_i P_i X P_i$是幂等的 Hermite 矩阵，且满足$\mathscr{R}(G_i) \subset \mathscr{R}(P_i)$，从而据习题30可证所需结论。

32. 证明上题的下列推论：若H为非负定的 Hermite 阵，而$X \in H\{2, 3, 4\}$，则X也是非负定的Hermite阵，且X的每个非

零特征值都是H的某特征值的倒数。

33. 给定 $A\in\mathbf{C}^{m\times n}$，证明

$$A\{2,3,4\}=\{YA^H \mid Y\in(A^HA)\{2,3,4\}\}$$

34. 试证：集合$A\{2,3,4\}$为有限集的充分必要条件是A^HA的非零特征值互异（即对应每个非零特征值的特征子空间是一维的）。此时，若有k个互异的非零特征值，则$A\{2,3,4\}$正好有2^k个元素。

35. 证明矩阵

$$A=\frac{1}{10}\begin{pmatrix} 9-3i & 12-4i & 10-10i \\ 3-3i & 4-4i & 0 \\ 6+6i & 8+8i & 0 \\ 6 & 8 & 0 \end{pmatrix}$$

恰好有下列4个$\{2,3,4\}$-逆

$$X_1=A^{+}=\frac{1}{70}\begin{pmatrix} 0 & 6+6i & 12-12i & 12 \\ 0 & 8+8i & 16-16i & 16 \\ 35+35i & -5-15i & -30+10i & -20-10i \end{pmatrix}$$

$$X_2=\frac{1}{60}\begin{pmatrix} -9-3i & 3+3i & 6-6i & 6 \\ -12-4i & 4+4i & 8-8i & 8 \\ 25+25i & -5-15i & -30+10i & -20-10i \end{pmatrix}$$

$$X_3=\frac{1}{420}\begin{pmatrix} 63+21i & 15+15i & 30-30i & 30 \\ 84+28i & 20+20i & 40-40i & 40 \\ 35+35i & 5+15i & 30-10i & 20+10i \end{pmatrix}$$

$$X_4=0$$

36. 证明：向量x是方程$Ax=b$的最小二乘解，其充分必要条件是存在向量r使得$\begin{bmatrix} r \\ x \end{bmatrix}$满足方程

$$\begin{bmatrix} I & A \\ A^H & 0 \end{bmatrix}\begin{bmatrix} r \\ x \end{bmatrix}=\begin{bmatrix} b \\ 0 \end{bmatrix}$$

37. 设A为习题24中的矩阵，又设$b=(-i,1,1)^T$。试证

$Ax=b$的最小二乘解的通式为

$$x=\frac{1}{19}\begin{pmatrix}0\\1\\0\\-4\\0\\0\end{pmatrix}+\begin{pmatrix}1&0&0&0&0&0\\0&0&-\frac{1}{2}&0&-1+2i&-\frac{1}{2}i\\0&0&1&0&0&0\\0&0&0&0&-2&-1-i\\0&0&0&0&1&0\\0&0&0&0&0&1\end{pmatrix}y$$

其中$y\in \mathbf{C}^6$是任意的。最小二乘的残差向量为

$$\frac{1}{19}(2i,\ 12,\ -2)^T$$

38． 证明上题中的最小范数最小二乘解为

$$x=\frac{1}{874}(0,\ 26-36i,\ 13-18i,\ -55-9i,\ -12-2i,$$

$$-46+59i)^T$$

39． 设$A\in \mathbf{C}^{m\times n}$，$b\in \mathbf{C}^m$，$a\in \mathbf{C}^n$。试证：当$Ax=b$有解时，使$\|x-a\|$达到极小的唯一解是

$$x=A^{(1,4)}b+(I-A^{(1,4)}A)a$$

40． 证明：对任意的$A\in \mathbf{C}^{m\times n}$，当$\lambda\in \mathbf{C}$通过复平面$\mathbf{C}$的原点的任一邻域趋向于零时，总有

$$\lim_{\lambda\to 0}(A^HA+\lambda I)^{-1}A^H=A^+$$

［证明］ 为此，只要证明对任意的$y\in \mathbf{C}^m$都有

$$\lim_{\lambda\to 0}(A^HA+\lambda I)^{-1}A^Hy=A^+y$$

就够了。由于$\mathscr{N}(A^H)=\mathscr{N}(A^+)$，所以对$y\in \mathscr{N}(A^H)$上式是成立的。故只须证上式对$y\in \mathscr{N}(A^H)^\perp=\mathscr{R}(A)$成立就行了。容易证明$A$是$\mathscr{R}(A^H)$到$\mathscr{R}(A)$上的1－1映射，所以对任意的$y\in \mathscr{R}(A)$，都有$x\in \mathscr{R}(A^H)$使得$y=Ax$。因此问题又等价于证明：对所有的$x\in \mathscr{R}(A^H)$皆有

$$\lim_{\lambda\to 0}(A^HA+\lambda I)^{-1}A^HAx=A^+Ax=x$$

成立，上面用到$P_{\mathscr{R}(A^H),\mathscr{N}(A)}=A^+A$。现设$\{u_1, u_2, \cdots, u_r\}$是$\mathscr{R}(A^H)$的基，且它们是由$A^HA$的特征向量构成，即满足

$$A^HAu_j=\lambda_ju_j,\ \lambda_j>0,\ j=1, 2, \cdots, r$$

于是有$x=\sum_{j=1}^{r}\beta_ju_j$。从而可进一步得到

$$(A^HA+\lambda I)^{-1}A^HAx=\sum_{j=1}^{r}\frac{\lambda_j\beta_j}{\lambda_j+\lambda}u_j,$$

$$-\lambda\neq\lambda_j,\ j=1, 2, \cdots, r$$

对上式取极限便得到

$$\lim_{\lambda\to 0}(A^HA+\lambda I)^{-1}A^HAx=\lim_{\lambda\to 0}\sum_{j=1}^{r}\frac{\lambda_j\beta_j}{\lambda_j+\lambda}u_j$$

$$=\sum_{j=1}^{r}\beta_ju_j=x$$

还可以通过证明等式

$$\lim_{\lambda\to 0}(A^HA+\lambda I)^{-1}A^HA=A^+A$$

$$=P_{\mathscr{R}(A^H),\mathscr{N}(A)}$$

来完成命题的证明。为此，设A^HA有满秩分解式：

$$A^HA=FF^H,\ F\in \mathbf{C}_r^{n\times r}$$

则等式

$$(A^HA+\lambda I)^{-1}A^HA=(FF^H+\lambda I)^{-1}FF^H$$

对一切λ（不是A^HA的特征值的相反数）成立。很容易证明有等式

$$(FF^H+\lambda I)^{-1}FF^H=F(F^HF+\lambda I_r)^{-1}F^H$$

成立。再注意F^HF是非奇异的，从而知

$$\lim_{\lambda\to 0}(F^H F+\lambda I_r)^{-1}=(F^H F)^{-1}$$

进而便得到

$$\lim_{\lambda\to 0}(A^H A+\lambda I)^{-1}A^H A=F(F^H F)^{-1}F^H$$

$$=FF^+=P_{\mathscr{R}(A^H),\mathscr{N}(A)}$$

上面最后一个等式成立是因为F的列向量为$\mathscr{R}(A^H A)=\mathscr{R}(A^H)$的一个基底（請思考一下，为什么？）。

41．设$A\in\mathbf{R}^{n\times n}$是对称的。指出$A\{1\}$中可以有不对称的，但也一定存在有对称的。

42．设$u\in\mathscr{R}(A)$，$v\in\mathscr{R}(A^T)$，且$v^T A^{(1)}u\neq -1$，$A\in\mathbf{R}^{n\times n}$。

试证 $\quad A^{(1)}-\dfrac{(A^{(1)}u)(v^T A^{(1)})}{1+v^T A^{(1)}u}\in(A+uv^T)\{1\}$。

43．若A是正规矩阵，试证：$A^+A=AA^+$。

44．设B，C，都是列满秩阵，且BC^T存在。证明：

$$(BC^T)^+=(C^T)^+B^+$$

45．证明每个方阵都有可逆的$\{1\}$-逆存在。

46．证明：若$\mathrm{rank}(A)=1$，则$A\{2,3,4\}=\{A^+,0\}$。

47．设$A\in\mathbf{C}^{m\times n}$。证明：极小化问题

$$\|AX-I\|_F=\min\text{且}\|X\|_F=\min$$

的解为$X=A^+$。也就是说矩阵方程$AX=I$的最小范数最小二乘解(按矩阵空间的Frobenius范数：$\|A\|_F=[\mathrm{tr}(A^H A)]^{\frac{1}{2}}$)为$A^+$。

第四章　矩阵分析

在数学的很多分支和工程实际中，特别涉及到多元分析时，仅仅研究矩阵代数是不够的，还必须研究常数矩阵序列和级数的收敛性，以及函数矩阵和矩阵函数的微积分，这便是矩阵分析的基本内容。本章内容的安排：先介绍矩阵的范数，讨论矩阵序列及级数的收敛，从而给出Lagrange-Sylvester定理；然后介绍矩阵函数和它的计算，以及在微分方程理论方面的应用；最后还给出函数矩阵的微分与积分的定义，以及某些矩阵的标量函数的微分和应用。

第一节　矩阵范数

在前面我们已经用到过向量或矩阵的范数，它是三维空间向量长度的概念的一种推广。虽然向量可作为一种特殊类型的矩阵统一处理，但矩阵作为线性变换的一种具体表示，有更为广泛的应用，所以其范数的定义更为特殊，需要稍为系统地介绍。后面将会看到矩阵范数在研究各种收敛性问题以及特征值的估计方面都特别有用。

我们知道，所有 $m\times n$ 阶复矩阵的全体，按通常的矩阵加法和矩阵与数的乘法规定了运算之后，便得到一个 mn维的复线性空间 $\mathbf{C}^{m\times n}$。要在这个纯代数结构的空间里研究矩阵序列和级数的收敛性，就必须引入求极限的运算。而极限运算是一个无限逼近的过程，矩阵的逼近，就是矩阵间的差异无限变小，而矩阵 A 与 B 的差异是用$A-B$ 的"大小" 来衡量的。这就需要对每个矩阵赋予一个体现其"大小"的量，即范数。实数的绝对值，复数的模，三

维空间向量的长度，都是抽象范数概念的原型。上述的三个对象统一记作x，衡量它们大小的量记为$\|x\|$，显然满足以下几个基本的事实：

(1) 正定性：$\|x\|\geqslant 0$，当且仅当$x=0$时$\|x\|=0$；

(2) 齐次性：$\|\lambda x\|=|\lambda|\,\|x\|$，$\lambda$为标量；

(3) 三角不等式：$\|x+y\|\leqslant\|x\|+\|y\|$。

事实上，有关向量"长度"的某些其它性质也可以从上述三条推导出来。所以数学上把这三条性质作为公理来定义抽象向量的范数。

定义4.1.1 若对任一$x\in \mathbf{C}^n$都有一个实数$\|x\|$与之对应，且满足

1. 正定性：$\|x\|\geqslant 0$，当且仅当$x=0$时$\|x\|=0$；
2. 齐次性：$\|\lambda x\|=|\lambda|\,\|x\|$，$\lambda\in\mathbf{C}$；
3. 三角不等式：$\|x+y\|\leqslant\|x\|+\|y\|$。

则称$\|x\|$为x的范数（norm）。定义了范数的$\mathbf{C}^n$又叫做一个线性赋范空间。

其实，在线性空间$\mathbf{C}^n$上赋范，就是在$\mathbf{C}^n$上定义一个满足上述三条性质的实函数$f(x)=\|x\|$。这种函数可以找到很多，确切地说有无穷多。例如，任取$x\in\mathbf{C}^n$且$x=(\xi_1,\ \xi_2,\ \cdots,\xi_n)^T$，我们就可定义

$$\|x\|_1=\sum_{i=1}^{n}|\xi_i|\quad,$$

$$\|x\|_2=\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{1/2}\quad,$$

$$\|x\|_\infty=\max_i|\xi_i|,$$

甚至

$$\|x\|_p=\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{1/p}\quad,\ 1\leqslant p<+\infty,$$

并且对于$\|x\|_1$，$\|x\|_\infty$很容易证明满足范数所要求的三条公

理。证明$\|x\|_2$，要用到第一章讲过的Cauchy-Schwarz 不等式：$\left|\sum_{i=1}^{n}\xi_i\eta_i\right|^2\leqslant \sum_{i=1}^{n}|\xi_i|^2\sum_{i=1}^{n}|\eta_i|^2$。而要证明$\|x\|_l$，则需要我们后面即将给出的Hölder不等式及Minkowski 不等式。在进行这些繁琐的推导之前，我们先指出几个重要的事实。

引理4.1.2 空间C^n上的任一种范数及x，$y\in C^n$，都有

$$|\|x\|-\|y\||\leqslant\|x-y\|$$

证明 据范数的第三条公理有

$$\|x\|=\|x-y+y\|\leqslant\|x-y\|+\|y\|$$
$$\|y\|=\|y-x+x\|\leqslant\|y-x\|+\|x\|$$

从上面两个关系式分别可推得

$$\|x\|-\|y\|\leqslant\|x-y\|$$
$$\|x\|-\|y\|\geqslant-\|x-y\|$$

综合起来便得

$$|\|x\|-\|y\||\leqslant\|x-y\|$$

定义4.1.3 对于x_m，$x\in C^n$，$m=1,2,\cdots$，若有

$$\lim_{m\to\infty}\|x_m-x\|=0$$

则称序列$\{x_m\}$以x为极限，记为

$$\lim_{m\to\infty}x_m=x$$

或$x_m\longrightarrow x$。当然，还应该强调在范数$\|\cdot\|$意义下，x_m收敛到x。

定理4.1.4 空间C^n上的任一种范数$\|\cdot\|$，作为C^n上的一个函数$f(x)=\|x\|$，都是连续的。

证明 任取$x\in C^n$及收敛到x的序列$\{x_m\}$。则据引理4.1.2有

$$|f(x_m)-f(x)|=|\|x_m\|-\|x\||\leqslant\|x_m-x\|\longrightarrow 0$$

故有$\lim_{m\to\infty}f(x_m)=f(x)$。从而证明了$f=\|\cdot\|$在$x$点连续。由于$x$是任取的，故$f$在$C^n$上连续。

这个定理表明，在$x_m \longrightarrow x$时，恒有

$$\lim_{m\to\infty} \| x_m \| = \| \lim_{m\to\infty} x_m \|$$

它是后面常用的一个结论。

既然在 $\mathbf{C}^n$ 上可以规定无穷多种范数，自然就会提出这样的问题：一个序列 $\{x_m\}$ 在不同的范数意义下同时都收敛或同时都发散吗？如果回答是肯定的，又会问：在处理问题时究竟采用哪种范数为宜？关于第一个问题的回答是：对于无穷维的线性空间，其上的各种范数未必是等价的。也就是说，同一个序列在不同的范数意义下，有的是收敛的，有的是发散的。而对于这里的空空间 $\mathbf{C}^n$（或其它有限维空间），由于其维数是有限的，所以不会出现这样的问题。对此下面将给出证明。至于第二个问题，也就是说各种范数都是等价的（这里暂指收敛与发散的一致性）情况下，如何选取范数的问题。依据是：既要考虑到实际问题的背景（是采用平方平均收敛，还是采用一致收敛，或是其它意义下的收敛），又要考虑到各种范数的内蕴性质给数学处理方面带来的可能性和简便性。

定义4.1.5 对线性空间X上的两种范数$\|\cdot\|_\alpha$，$\|\cdot\|_\beta$，如果存在正数k_1，k_2，使得对一切 $x\in X$都有

$$k_1 \| x \|_\beta \leqslant \| x \|_\alpha \leqslant k_2 \| x \|_\beta \qquad (*)$$

则称这两种范数是等价的。

从定义可以看出，如果序列 $\{x_m\}$ 在范数$\|\cdot\|_\alpha$下是收敛的，则由(*)的左半部分得知在范数$\|\cdot\|_\beta$下也是收敛的；如果序列 $\{x_m\}$ 在范数$\|\cdot\|_\alpha$下是发散的，则由(*)的右半部分得知在范数$\|\cdot\|_\beta$下也是发散的。为了证明有限维空间X上的任何两种范数互相等价，这里不加证明的引用一个结论，有兴趣的读者可参阅Erwin Kreyszig的《Introductory Functional Analysis with Applications》中引理 2.4.1（有中译本，北京航空学院出版社，1987.）。

引理4.1.6（线性组合） 设X是一个线性赋范空间（不必是有限维的），$\{x_1, x_2, \cdots, x_n\}$是$X$中的一组线性无关的向量。则对任意给定的一组标量$\alpha_1, \alpha_2, \cdots, \alpha_n$，都存在一个正数$c$，使得

$$\|\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n\| \geqslant c \sum_{i=1}^{n} |\alpha_i|$$

这个引理的证明要用到数学分析中Bolzano-Weierstrass定理或与之等价的其它定理，叙述起来较繁琐，故略。现用它来证明

定理4.1.7 在有限维线性空间X上的所有范数都互相等价。

证明 设$\dim X = n$，$\{e_1, e_2, \cdots, e_n\}$是$X$的一个基底，$\|\cdot\|_\alpha$，$\|\cdot\|_\beta$为$X$上的任意两种范数。则任取$x \in X$，有

$$x = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n$$

据引理4.1.6，存在常数$c > 0$，使得

$$\|x\|_\alpha \geqslant c \sum_{i=1}^{n} |\alpha_i| \quad ,$$

即有

$$\sum_{i=1}^{n} |\alpha_i| \leqslant \frac{1}{c} \|x\|_\alpha$$

另外，据三角不等式还有

$$\|x\|_\beta = \|\sum_{i=1}^{n} \alpha_i e_i\|_\beta \leqslant \sum_{i=1}^{n} |\alpha_i| \, \|e_i\|_\beta \leqslant k \sum_{i=1}^{n} |\alpha_i|$$

其中 $k = \max_i \|e_i\|_\beta$。从而有

$$\|x\|_\beta \leqslant k \sum_{i=1}^{n} |\alpha_i| \leqslant \frac{k}{c} \|x\|_\alpha$$

取$k_1 = c/k$，便得到

$$k_1 \|x\|_\beta \leqslant \|x\|_\alpha$$

在上述证明中，把$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$的位置调换一下，便可得到

$$\| x \|_\alpha \leqslant k_2 \| x \|_\beta$$

综合起来，便有

$$k_1 \| x \|_\beta \leqslant \| x \|_\alpha \leqslant k_2 \| x \|_\beta$$

从而证明了定理。

现在回到两个重要不等式的证明上。为此，先给出

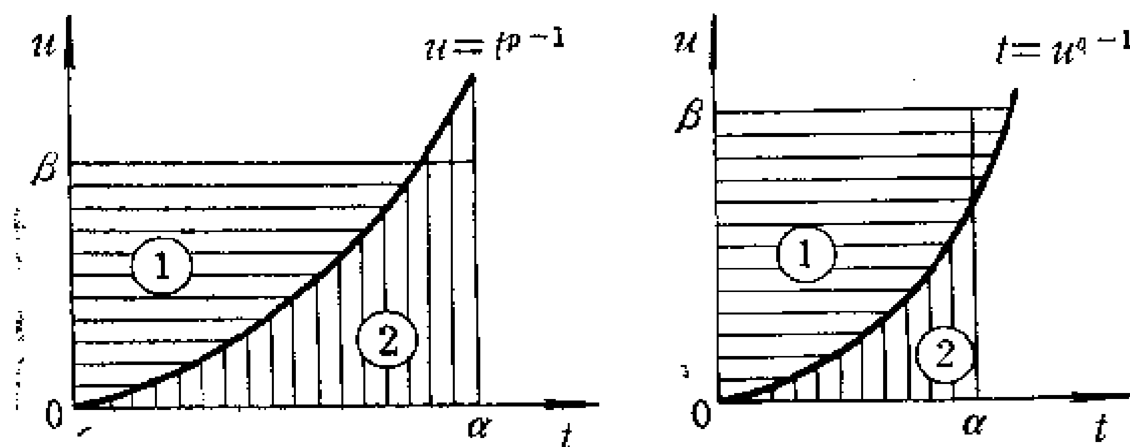
引理4.1.8 若正实数 p，q互异且满足

$$\frac{1}{p} + \frac{1}{q} = 1$$

则对任何非负实数α，β总有

$$\alpha \cdot \beta \leqslant \frac{\alpha^p}{p} + \frac{\beta^q}{q}$$

证明 由于$p \neq q$，故可设$p = 1 + \mu$，$q = 1 + \frac{1}{\mu}$，$\mu > 0$，如



图所示，在u-t平面上作出函数曲线

$$u = t^\mu, \qquad 0 \leqslant t \leqslant \alpha, \ 0 \leqslant u \leqslant \beta$$

计算图上所标明的面积：

$$\text{面积①} = \int_0^\beta u^{\frac{1}{\mu}} \, du = \frac{\beta^q}{q}$$

$$\text{面积②} = \int_0^\alpha t^\mu \, dt = \frac{\alpha^p}{p}$$

很显然有

$$\text{面积①} + \text{面积②} \geqslant \alpha \cdot \beta$$

定理4.1.9 （Hölder）对任意的x，$y \in \mathbf{C}^n$及引理4.1.8中所给定的p，q，当$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$，$y=(\eta_1, \eta_2, \cdots, \eta_n)^T$时，总有

$$\sum_{i=1}^{n} |\xi_i \eta_i| \leqslant \left(\sum_{i=1}^{n} |\xi_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |\eta_i|^q \right)^{1/q}$$

证明 令

$$\alpha = \left(\sum_{i=1}^{n} |\xi_i|^p \right)^{1/p}, \qquad \beta = \left(\sum_{i=1}^{n} |\eta_i|^q \right)^{1/q}$$

显然，若$\alpha \cdot \beta = 0$，不等式是成立的。现就$\alpha \cdot \beta \neq 0$来证。置

$$\gamma_k = |\xi_k| / \alpha, \qquad \delta_k = |\eta_k| / \beta$$

据引理4.1.8有

$$\gamma_k \delta_k \leqslant \frac{\gamma_k^p}{p} + \frac{\delta_k^q}{q}, \qquad k=1, 2, \cdots, n$$

上式两端关于k从1到n取和，便有

$$\frac{1}{\alpha\beta} \sum_{k=1}^{n} |\xi_k \eta_k|$$

$$= \sum_{k=1}^{n} \gamma_k \delta_k \leqslant \frac{1}{p} \sum_{k=1}^{n} \gamma_k^p + \frac{1}{q} \sum_{k=1}^{n} \delta_k^q$$

$$= \frac{1}{p} + \frac{1}{q} = 1$$

从而有

$$\sum_{k=1}^{n} |\xi_k \eta_k| \leqslant \alpha\beta = \left(\sum_{k=1}^{n} |\xi_k|^p \right)^{1/p} \left(\sum_{k=1}^{n} |\eta_k|^q \right)^{1/q}$$

在Hölder不等式中，取$p=q=2$，便是著名的Cauchy-Schwarz不等式。

定理4.1.10 （Minkowski）对任意的x，$y \in \mathbf{C}^n$且$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$，$y=(\eta_1, \eta_2, \cdots, \eta_n)^T$，$p \geqslant 1$，则有

$$\left(\sum_{i=1}^{n}|\xi_i+\eta_i|^p\right)^{1/p}\leqslant\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{1/p}+\left(\sum_{i=1}^{n}|\eta_i|^p\right)^{1/p}$$

证明 $p=1$时，不等式显然成立，故设$p>1$。为简化推导，记$\omega_j=\xi_j+\eta_j$，从而由三角不等式可得

$$\begin{aligned}|\omega_j|^p&=|\xi_j+\eta_j||\omega_j|^{p-1}\\&\leqslant(|\xi_j|+|\eta_j|)|\omega_j|^{p-1}\end{aligned}$$

上式两端关于j求和，又有

$$\sum|\omega_j|^p\leqslant\sum|\xi_j||\omega_j|^{p-1}+\sum|\eta_j||\omega_j|^{p-1}$$

上式右端的两个和式分别用Hölder不等式可得

$$\sum|\xi_j||\omega_j|^{p-1}\leqslant\left(\sum|\xi_j|^p\right)^{1/p}\left[\sum(|\omega_j|^{p-1})^q\right]^{1/q}$$

$$\sum|\eta_j||\omega_j|^{p-1}\leqslant\left(\sum|\eta_j|^p\right)^{1/p}\left[\sum(|\omega_j|^{p-1})^q\right]^{1/q}$$

注意到$(p-1)q=p$，上面三式结合起来便有

$$\sum|\omega_j|^p\leqslant\left[\left(\sum|\xi_j|^p\right)^{1/p}+\left(\sum|\eta_j|^p\right)^{1/p}\right]\left(\sum|\omega_j|^p\right)^{1/q}$$

上面不等式两端除上$\left(\sum|\omega_j|^p\right)^{1/q}$，注意到$1-1/q=1/p$，便得到 Minkowski 不等式。

有了上述不等式，便很容易证明下式

$$\|x\|_p=\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{1/p}\quad 1\leqslant p<+\infty$$

确实在$\mathbf{C}^n$上定义了各种范数，通称p-范数。請读者自己证明：对任意的x，$y\in\mathbf{C}^n$有

$$\|x+y\|_p\leqslant\|x\|_p+\|y\|_p$$

$$\lim_{p\to\infty}\|x\|_p=\|x\|_\infty$$

前面讨论了空间$\mathbf{C}^n$上的范数。由于每一个$m\times n$阶矩阵，都可以看作为空间 $\mathbf{C}^{mn}$ 中的一个向量，所以对矩阵 $A=(a_{ij})\in\mathbf{C}^{m\times n}$，完全可以用同样的公理来定义其范数：

定义4.1.11 若对任一 $A\in\mathbf{C}^{m\times n}$，都有实数 $\|A\|$ 与之对应，且满足

1. 正定性：$\|A\|\geqslant0$，当且仅当$A=0$时，$\|A\|=0$；
2. 齐次性：$\|\lambda A\|=|\lambda|\,\|A\|$，$\forall\lambda\in\mathbf{C}$；

3. 三角不等式：$\| A+B \| \leqslant \| A \| + \| B \|$。

则称 $\| A \|$ 为A的范数。

按照上面的定义，与向量 $x \in \mathbf{C}^n$ 的几种范数相对应，矩阵 $A=(a_{ij}) \in \mathbf{C}^{m \times n}$ 有范数

$$\| A \| = \sum_{j=1}^{m} \sum_{j=1}^{n} |a_{ij}|, \qquad \| A \| = \max_{i,j} |a_{ij}|,$$

$$\| A \| = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}, \quad p \geqslant 1$$

这里在范数符号 $\| \cdot \|$ 中没有使用下脚，主要是留下来为后面使用，不是疏忽。上面的范数是互相等价的，所以在定性地研究收敛性问题时，都可以采用，而且以 $\| A \| = \max\limits_{i,j} |a_{ij}|$ 最为简洁。

但通常更喜欢用 $\| A \| = \left(\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$，它叫做 Frobenius 范数，故又记作为 $\| A \|_F$。因为这种范数就是酉空间 $(\mathbf{C}^{m \times n}, (A|B)=\operatorname{tr}(B^H A))$ 中的内积所诱导的范数：

$$\| A \|_F^2 = (A|A) = \operatorname{tr}(A^H A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2$$

它有一系列的良好性质，有明显的几何意义，所以在矩阵的最小二乘逼近时常用它。下面只指出两个最重要的特性。

定理4.1.12 $\| \cdot \|_F$ 是酉不变的。也就是说，对酉矩阵 $U \in \mathbf{C}^{m \times m}$，$V \in \mathbf{C}^{n \times n}$，恒有

$$\| A \|_F = \| UA \|_F = \| AV \|_F = \| UAV \|_F$$

证明 利用等式 $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ 及 $U^H U = I_m$，$VV^H = I_n$ 立即可得

$$\| UA \|_F^2 = \operatorname{tr}(A^H U^H U A) = \operatorname{tr}(A^H A) = \| A \|_F^2$$

$$\begin{aligned} \| AV \|_F^2 &= \operatorname{tr}(AVV^H A^H) = \operatorname{tr}(AA^H) \\ &= \operatorname{tr}(A^H A) = \| A \|_F^2 \end{aligned}$$

$$\begin{aligned} \| UAV \|_F^2 &= \operatorname{tr}(V^H A^H U^H U A V) \\ &= \operatorname{tr}(V^H A^H A V) = \operatorname{tr}(A^H A) = \| A \|_F^2 \end{aligned}$$

定理4.1.13 设$A\in \mathbf{C}^{m\times n}$，$B\in \mathbf{C}^{n\times l}$，则有

$$\|AB\|_F\leqslant \|A\|_F\|B\|_F$$

证明 对A和B分别作行分块和列分块：

$$A=\begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix} \qquad B=[b_1, b_2, \cdots, b_l]$$

其中a_i，$b_j\in \mathbf{C}^n$。据$\|\cdot\|_F$的定义知

$$\|A\|_F^2=\sum_{i=1}^m \|a_i\|_2^2 \qquad \|B\|_F^2=\sum_{j=1}^l \|b_j\|_2^2$$

$$\|AB\|_F^2=\sum_{i=1}^m \sum_{j=1}^l |(AB)_{ij}|^2$$

其中$(AB)_{ij}=a_i^Tb_j$，据Hölder不等式有

$$|a_i^Tb_j|^2\leqslant \|a_i\|_2^2\|b_j\|_2^2$$

从而便得到

$$\|AB\|_F^2=\sum_{i=1}^m \sum_{j=1}^l |a_i^Tb_j|^2\leqslant \sum_{i=1}^m \sum_{j=1}^l \|a_i\|_2^2\|b_j\|_2^2$$
$$=\|A\|_F^2\|B\|_F^2$$

此不等式两端开方便为

$$\|AB\|_F\leqslant \|A\|_F\|B\|_F$$

上述定理给出的F-范数的性质，不是所有矩阵范数都具备的。比如，按$\|A\|=\max\limits_{i,j}|a_{ij}|$定义的范数，当取

$$A=B=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

时，可以验证不满足$\|AB\|\leqslant\|A\|\|B\|$。特别是当$A\in \mathbf{C}^{n\times n}$，把$A$视为从$\mathbf{C}^n$到$\mathbf{C}^n$的线性变换，而把定理4.1.13中的$B$换成特殊的$x\in \mathbf{C}^n$时，则有

$$\|x\|_F=\left(\sum_{i=1}^n |\xi_i|^2\right)^{1/2}=\|x\|_2$$

及 $$\|Ax\|_2\leqslant\|A\|_F\|x\|_2$$
这种不等式的成立，在分析的研究中是十分重要的。所以定义矩阵的范数不能和定义向量的范数完全相同，应该根据特殊的需要附加一定的条件。下面给出相容的矩阵范数的定义：

定义4.1.14 在矩阵空间$\mathbf{C}^{n\times n}$上的任一实函数，记为$\|\cdot\|$，如果对所有的A，$B\in\mathbf{C}^{n\times n}$，$\lambda\in\mathbf{C}$都满足

1. $\|A\|\geqslant 0$，当且仅当$A=0$时有$\|A\|=0$；
2. $\|\lambda A\|=|\lambda|\,\|A\|$；
3. $\|A+B\|\leqslant\|A\|+\|B\|$；
4. $\|AB\|\leqslant\|A\|\,\|B\|$。

则称$\|\cdot\|$为相容的矩阵范数，或简称矩阵范数。

定义4.1.15 如果一个矩阵范数$\|A\|$和一个向量范数$\|x\|$对一切向量x和矩阵A都满足
$$\|Ax\|\leqslant\|A\|\,\|x\|$$
则称这两个范数是相容的。

上面两个定义虽然都有相容的字样，但是要注意区别。根据定义，矩阵的F-范数是一种相容的矩阵范数，并且这种矩阵范数与向量x的欧氏范数（$\|x\|_2$）是相容的。同时可以证明，任意给定一种矩阵范数$\|A\|_m$，一定可以找到一个与之相容的向量范数$\|x\|_v$。比如，任取向量$a\neq 0$，令$\|x\|_v=\|xa^H\|_m$，则很容易验证它确实定义了一种向量范数。这种向量范数与矩阵范数相容，是因为
$$\begin{aligned}\|Ax\|_v&=\|Axa^T\|_m\leqslant\|A\|_m\|xa^T\|_m\\&=\|A\|_m\|x\|_v\end{aligned}$$
这里范数$\|\cdot\|$的下脚m和v，分别表示矩阵（matrix）范数与向量（vector）范数的意思，以示区别。最常用的与向量范数相容的矩阵范数是所谓的算子范数：
$$\|A\|=\sup_{\|x\|=1}\|Ax\|$$
其中向量的范数$\|x\|$与$\|Ax\|$可以相同，也可以不同。不管

怎样，总有$\|Ax\| \leqslant \|A\| \|x\|$成立。矩阵的算子范数的计算归结为求函数的约束极值。当向量的范数选得适当时，是有可能计算出来的。一般情况下，从分析的角度来看一定存在（连续函数在有界闭集上可达到极大极小值），但计算不是很容易的。下面给出几种特殊情形的算例。

例 对于$A \in \mathbf{C}^{m \times n}$，当向量$x$，$Ax$的范数都分别取$\|\cdot\|_1$，$\|\cdot\|_\infty$，$\|\cdot\|_2$时，其对应的算子范数也记为$\|A\|_1$，$\|A\|_\infty$，$\|A\|_2$，则有

1. $\|A\|_1 = \sup_{\|x\|_1=1} \|Ax\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$

2. $\|A\|_\infty = \sup_{\|x\|_\infty=1} \|Ax\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$

3. $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2 = [\rho(A^H A)]^{1/2} = \sigma_1$

其中σ_1为A的最大奇异值。

证明 任取$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$且$\|x\|_1 = \sum_{i=1}^{n} |\xi_i| = 1$，则有

$$Ax = \left(\sum_{k=1}^{n} a_{1k}\xi_k, \sum_{k=1}^{n} a_{2k}\xi_k, \cdots, \sum_{k=1}^{n} a_{nk}\xi_k\right)^T$$

$$\|Ax\|_1 = \sum_{l=1}^{m} \left|\sum_{k=1}^{n} a_{lk}\xi_k\right|$$

$$\leqslant \sum_{l=1}^{m} \sum_{k=1}^{n} |a_{lk}||\xi_k| = \sum_{k=1}^{n}\left(|\xi_k| \sum_{l=1}^{m} |a_{lk}|\right)$$

$$\leqslant \left(\max_k \sum_{l=1}^{m} |a_{lk}|\right) \sum_{k=1}^{n} |\xi_k| = \max_k \sum_{l=1}^{m} |a_{lk}|$$

由于x的任意性，从而证明了

$$\sup_{\|x\|_1=1} \|Ax\|_1 \leqslant \max_k \sum_{l=1}^{m} |a_{lk}|$$

如果 $\sum_{i=1}^{m}|a_{is}|=\max_{k}\sum_{i=1}^{m}|a_{ik}|$，则取 $x=e_s$，便有

$$\|Ae_s\|_1=\|A_s\|_1=\sum_{i=1}^{m}|a_{is}|$$

其中 A_s 表示矩阵A的第 s 个列向量。从而说明函数 $\|Ax\|_1$ 在单位球面 $\|x\|_1=1$上的极大值点为e_s。这就证明了

$$\|A\|_1=\sup_{\|x\|_1=1}\|Ax\|_1=\max_{j}\sum_{i=1}^{m}|a_{ij}|$$

同样的推理可证 $\|A\|_\infty=\sup_{\|x\|_\infty=1}\|Ax\|_\infty=\max_{i}\sum_{j=1}^{n}|a_{ij}|$。

至于，$\|A\|_2=\sup_{\|x\|_2=1}\|Ax\|_2=[\rho(A^HA)]^{1/2}$ 利用酉空间内积与范数的关系及正规阵酉相似于对角阵的结论便可推出。任取 $x=(\xi_1, \xi_2, \cdots, \xi_n)^T$ 且 $\|x\|_2=1$。则因 $\|Ax\|_2^2=(Ax|Ax)=x^HA^HAx$，故设正规阵 A^HA 的特征值（即 A 的奇异值之平方）为

$$\sigma_1^2\geqslant\sigma_2^2\geqslant\cdots\geqslant\sigma_n^2\geqslant0$$

对应的标准正交特征向量为

$$\{u_1, u_2, \cdots, u_n\}$$

即有 $A^HAu_i=\sigma_i^2u_i$，$(u_i|u_j)=u_j^Hu_i=\delta_{ij}$。作酉阵

$$U=(u_1, u_2, \cdots, u_n)$$

便有$A^HA=UDU^H$，其中$D=\mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$。从而有

$$\|Ax\|_2^2=x^HUDU^Hx$$

令$y=U^Hx$，则 $\|y\|_2^2=\|U^Hx\|_2^2=\|x\|_2^2=1$。故而可得

$$\|Ax\|_2^2=y^HDy=\sum_{i=1}^{n}\sigma_i^2|\eta_i|^2\leqslant\sigma_1^2$$

其中用到 $y^Hy=\sum_{i=1}^{n}|\eta_i|^2=1$。这样便由 $\|Ax\|_2\leqslant\sigma_1$ 及 x 的任意性推得

$$\|A\|_2=\sup_{\|x\|_2=1}\|Ax\|_2\leqslant\sigma_1$$

特别取$x=u_1$（显然$\|x\|_2=\|u_1\|_2=1$），又有

$$\|Au_1\|_2^2=u_1^HA^HAu_1=\sigma_1^2u_1^Hu_1=\sigma_1^2$$

即$\|Au_1\|_2=\sigma_1$，这说明$\|Ax\|_2$在单位球面$\|x\|_2=1$上有极大值点为u_1，从而证明了

$$\|A\|_2=\sup_{\|x\|_2=1}\|Ax\|_2=\sigma_1=[\rho(A^HA)]^{1/2}$$

在矩阵A的算子范数中，$\|A\|_2$是最常用的。因为我们经常在酉空间（或欧氏空间）考虑问题，而这种范数与向量的欧氏范数是相容的。另外，它还有如下的性质：

定理4.1.16 对任一$A\in\mathbf{C}^{m\times n}$，都有

1. $\|A\|_2=\max\{|y^HAx|\mid\|x\|_2=\|y\|_2=1\}$
2. $\|A\|_2=\|\overline{A}\|_2=\|A^T\|_2=\|A^H\|_2$
3. $\|A^HA\|_2=\|A\|_2^2$
4. 对于酉阵U，V有$\|A\|_2=\|UAV\|_2$。

证明 1. 任取$x\in\mathbf{C}^n$，$y\in\mathbf{C}^m$且$\|x\|_2=\|y\|_2=1$，则由Cauchy-Schwarz 不等式可得

$$|y^HAx|=|(Ax|y)|\leqslant\|y\|_2\|Ax\|_2$$
$$\leqslant\|y\|_2\|A\|_2\|x\|_2=\|A\|_2$$

据x、y的任意性，便得

$$\max\{|y^HAx|\mid\|x\|_2=\|y\|_2=1\}\leqslant\|A\|_2$$

现证明上式中的等号成立，也就是说要证明存在x_0与y_0且$\|x_0\|_2=\|y_0\|_2=1$，使得$|y_0^HAx_0|=\|A\|_2$。当$A=0$时，1是自然成立的。为此，设$A\neq0$，则存在x_0满足$\|x_0\|_2=1$且$\|Ax_0\|_2=\|A\|_2\neq0$。现取

$$y_0=Ax_0/\|Ax_0\|_2$$

显然$\|y_0\|_2=1$，并且有

$$|y_0^HAx_0|=|(Ax_0|y_0)|$$
$$=|(Ax_0|Ax_0/\|Ax_0\|_2)|$$

$$= (Ax_0 | Ax_0) / \| Ax_0 \|_2$$
$$= \| Ax_0 \|_2 = \| A \|_2$$

从而证明了1。

2的证明根据 $e(A^H A) = e(AA^H)$ 及 $|f(z)| = |f(\bar{z})|$ 很容易推出。其中 $\rho(A)$ 表示 A 的谱半径，$f(z)$ 表示复值函数，这里用到的 $f(z) = z^H Bz$，而 B 为Hermite阵，$z \in \mathbf{C}^n$（或 $\mathbf{C}^m$）。

3的证明与前面例子中的3的证明类同。

4的证明根据3的结果可以推出：

$$\| A \|_2^2 = \| A^H A \|_2 = \| A^H U^H U A \|_2$$
$$= \| UA \|_2^2 = \| A^H U^H \|_2^2$$
$$= \| V^H A^H U^H \|_2^2 = \| UAV \|_2^2$$

关于矩阵的范数就介绍到这里。最后再强调一下，在矩阵空间 $\mathbf{C}^{m \times n}$ 上的各种范数都是互相等价的。因而在考虑收敛性问题时，取哪一种范数，所得收敛或发散的结论是一致的。

第二节 矩阵序列与矩阵级数

设 A_1，A_2，…，A_k，…是空间 $\mathbf{C}^{m \times n}$ 中的无穷矩阵序列，记之为 $\{A_k\}$，$(A_k)_{ij}$ 表示矩阵 A_k 的第 (i, j) 元素。如果有矩阵 $A = (a_{ij}) \in \mathbf{C}^{m \times n}$，使得

$$\lim_{k \to \infty} (A_k)_{ij} = a_{ij}, \qquad i = 1, 2, \cdots, m, \ j = 1, 2, \cdots, n$$

成立，则称矩阵序列 $\{A_k\}$ 以 A 为极限，或 $\{A_k\}$ 收敛到 A，记为

$$\lim_{k \to \infty} A_k = A$$

如此看来，一个矩阵序列 $\{A_k\} \subset \mathbf{C}^{m \times n}$ 的收敛等价于 $m \cdot n$ 个标量序列 $\{(A_k)_{ij}\}$ $(1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n)$ 的同时收敛。因此，可以用初等分析的方法来研究它。但考虑到要同时研究 $m \cdot n$ 个标量序列，未免繁琐。因此我们用矩阵的范数作为工具来研究

它。首先给出定义

定义4.2.1 对于矩阵序列 $A_k \in \mathbf{C}^{m\times n}$，$k=1, 2, \cdots$，如果存在矩阵 $A \in \mathbf{C}^{m\times n}$，使得

$$\lim_{k\to\infty} \| A_k - A \| = 0$$

则称矩阵序列 $\{A_k\}$ 收敛到 A，记为 $\lim\limits_{k\to\infty} A_k = A$。其中 $\|\cdot\|$ 为空间 $\mathbf{C}^{m\times n}$ 上的任一种范数。

这里自然会提出这样的问题：两种定义是否等价？下面给出证明。由于有限维空间上的各种范数互相等价，所以只要取一种范数证明就够了。

定理4.2.2 对于 A_k，$A \in \mathbf{C}^{m\times n}$；$k=1, 2, \cdots$，总有

$$\| A_k - A \|_F \longrightarrow 0 \Longleftrightarrow |(A_k)_{ij} - a_{ij}| \longrightarrow 0,$$
$$1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n,$$

证明 据 F-范数的定义有

$$\| A_k - A \|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |(A_k)_{ij} - a_{ij})^2$$

显然有

$$\| A_k - A \|_F \longrightarrow 0 \Longleftrightarrow |(A_k)_{ij} - a_{ij}| \longrightarrow 0,$$
$$1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n。$$

如同标量序列的性质一样，对收敛的矩阵序列也有一系列的性质。

定理4.2.3 收敛的矩阵序列 $\{A_k\}$ 是有界的，也就是说，存在正数 M 使得对一切 k 都有

$$\| A_k \| \leqslant M$$

证明 设 $\{A_k\}$ 收敛到 A，即

$$\lim_{k\to\infty} \| A_k - A \| = 0$$

据标量序列 $\delta_k = \| A_k - A \| \longrightarrow 0$ 的定义，任给定 $\varepsilon_0 > 0$，都存在 N，使得当 $k > N$ 时有 $\delta_k < \varepsilon_0$，即有

$$\| A_k - A \| < \varepsilon_0, \qquad k = N+1, N+2, \cdots,$$

从而据三角不等式有

$$\| A_k \| = \| A_k - A + A \| \leqslant \| A_k - A \| + \| A \|$$
$$< \| A \| + \varepsilon_0, \quad k \geqslant N+1,$$

再令 $K = \max\{ \| A_1 \|, \| A_2 \|, \cdots, \| A_N \| \}$，并取

$$M = \max(K, \| A \| + \varepsilon_0)$$

则对一切k，都有$\| A_k \| \leqslant M$。

定理4.2.4 若 $A_k, B_k \in \mathbf{C}^{n\times n}$，$\alpha_k, \beta_k \in \mathbf{C}$，且有$\lim\limits_{k\to\infty} A_k = A$，$\lim\limits_{k\to\infty} B_k = B$，$\lim\limits_{k\to\infty} \alpha_k = \alpha$，$\lim\limits_{k\to\infty} \beta_k = \beta$ 则有

1. $\lim\limits_{k\to\infty} (\alpha_k A_k + \beta_k B_k) = \alpha A + \beta B$
2. $\lim\limits_{k\to\infty} A_k B_k = AB$
3. 若 A_k^{-1}，A^{-1}存在，则$\lim\limits_{k\to\infty} A_k^{-1} = A^{-1}$

证明 为了方便，我们采用相容的矩阵范数，也就是说，对于任意的A，$B \in \mathbf{C}^{n\times n}$，总有

$$\| AB \| \leqslant \| A \| \, \| B \|$$

再利用三角不等式及收敛序列的有界性便可得到

$$\begin{aligned}
&\| \alpha_k A_k + \beta_k B_k - \alpha A - \beta B \| \\
&= \| \alpha_k A_k - \alpha A \| + \| \beta_k B_k - \beta B \| \\
&\leqslant \| \alpha_k A_k - \alpha_k A + \alpha_k A - \alpha A \| \\
&\quad + \| \beta_k B_k - \beta_k B + \beta_k B - \beta B \| \\
&\leqslant |\alpha_k| \, \| A_k - A \| + |\alpha_k - \alpha| \, \| A \| + \\
&\quad |\beta_k| \, \| \beta_k - B \| + |\beta_k - \beta| \, \| B \| \longrightarrow 0
\end{aligned}$$

$$\| A_k B_k - AB \| = \| A_k B_k - AB_k + AB_k - AB \|$$
$$\leqslant \| A_k - A \| \, \| B_k \| + \| A \| \, \| B_k - B \| \longrightarrow 0$$

这样便证明了1、2。至于3，考虑矩阵序列

$$A_k A_k^{-1} = I, \qquad k = 1, 2, \cdots$$

利用已证明的2，便可获证。

根据上述定理易知，当$\{ A_k \}$收敛时，则对任意的P，Q

都有$\lim_{k\to\infty} PA_kQ = P\lim_{k\to\infty} A_kQ$。特别是当矩阵序列$\{A_k\}$的一般项 $A_k = B^k$,即为方阵B的k次幂时,这是后面要重点研究的一种矩阵序列。如果序列$\{B^k\}$是收敛的,则称矩阵B是幂收敛的。很显然与B相似的矩阵PBP^{-1}也是幂收敛的。又当B是块对角阵时,即

$$B = \mathrm{diag}(B_1, B_2, \cdots, B_s)$$

由于有

$$B^k = \mathrm{diag}(B_1^k, B_2^k, \cdots, B_s^k)$$

所以$\{B^k\}$收敛的充分必要条件是$\{B_i^k\}$ $i=1,2,\cdots s$,都收敛。下面我们来考查方阵$A\in \mathbf{C}^{n\times n}$幂收敛的充分必要条件是什么。

我们知道每个$A\in \mathbf{C}^{n\times n}$都有一个与之相似的Jordan标准形$J$,即存在满秩阵$P$使得

$$P^{-1}AP = J = \mathrm{diag}(J_1, J_2, \cdots, J_k)$$

其中

$$J_i = J_{n_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix} \in \mathbf{C}^{n_i\times n_i}$$

并且$n = \sum_{i=1}^{k} n_i$。我们把$J_{n_i}(\lambda_i)$分解为

$$J_{n_i}(\lambda_i) = \lambda_i I_{n_i} + U_{n_i}$$

其中

$$U_{n_i} = J_{n_i}(0) = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix} \in \mathbf{C}^{n_i\times n_i}$$

这个矩阵有一个很好的性质：U_{n_i}的幂次每增加1，主对角线上方的一排1就往右上方平移一次，特别有

$$U_{n_i}^{n_i-1}=\begin{pmatrix} 0 & 0 & \cdots & \cdots & 1 \\ & 0 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & 0 \\ & & & & 0 \end{pmatrix},\quad U_{n_i}^{l}=O,\quad l\geqslant n_i$$

这就给我们以很大的方便。根据二项式定理有

$$\begin{aligned} J_{n_i}^{l}(\lambda_i)&=[\lambda_i I+U_{n_i}]^{l} \\ &=\lambda_i^{l}I_{n_i}+l\lambda_i^{l-1}U_{n_i}+\frac{l(l-1)}{2}\lambda_i^{l-2}U_{n_i}^{2}+ \\ &\quad\cdots+C_l^{j}\lambda_i^{l-j}U_{n_i}^{j}+\cdots+U_{n_i}^{l} \end{aligned}$$

特别当 $l\geqslant n_i$时，$U_{n_i}^{l}=O$，故这时上式为

$$\begin{aligned} J_{n_i}^{l}(\lambda_i)&=\lambda_i^{l}I_{n_i}+C_l^{1}\lambda_i^{l-1}U_{n_i}+C_l^{2}\lambda_i^{l-2}U_{n_i}^{2}+\cdots+ \\ &\quad+C_l^{j}\lambda_i^{l-j}U_{n_i}^{j}+\cdots+C_l^{n_i-1}\lambda_i^{l-n_i+1}\ U_{n_i}^{n_i-1} \end{aligned}$$

$$=\begin{pmatrix} \lambda_i^{l} & C_l^{1}\lambda_i^{l-1} & C_l^{2}\lambda_i^{l-2} & \cdots & C_l^{n_i-2}\lambda_i^{l-n_i+2} & C_l^{n_i-1}\lambda_i^{l-n_i+1} \\ & \lambda_i^{l} & C_l^{1}\lambda_i^{l-1} & \cdots & \vdots & C_l^{n_i-2}\lambda_i^{l-n_i+2} \\ & & \lambda_i^{l} & & \vdots & \\ & & & \ddots & \vdots & \\ & & & & \lambda_i^{l} & C_l^{1}\lambda_i^{l-1} \\ & & & & & \lambda_i^{l} \end{pmatrix}$$

或者写成更紧凑的形式

$$J_{n_i}^{l}(\lambda_i)=\sum_{j=0}^{n_i-1}C_l^{j}\lambda_i^{l-j}U_{n_i}^{j}\quad,\quad l\geqslant n_i$$

这样，我们就可得到Jordan块$J_{n_i}(\lambda_i)$幂收敛的充分必要条件。

引理4.2.5 Jordan块 $J_{n_i}(\lambda_i)$ 是幂收敛的，当且仅当 $|\lambda_i|<1$ 或 $\lambda_i=1$ 且 $n_i=1$（这时 $J_{n_i}(\lambda_i)$ 为一阶的矩阵）。

证明 若 J_{n_i} 是幂收敛的，则据收敛的定义可知 $\lim\limits_{l\to\infty}(\lambda_i^l)$ 存在，因此有 $|\lambda_i|\leqslant 1$。当 $|\lambda_i|\neq 1$ 时，J_{n_i} 收敛到零矩阵 O_{n_i}。当 $|\lambda_i|=1$，对下式两端取极限：

$$(\lim_{k\to\infty}\lambda_i^k)\lambda_i=\lim_{k\to\infty}\lambda_i^{k+1}$$

便可得到 $\lambda_i=1$。这时要 $J_{n_i}(\lambda_i)$ 幂收敛，则必须 $n_i=1$，即 $J_{n_i}(\lambda_i)=[1]$，否则在 $J_{n_i}^l(\lambda_i)$ 中有元素 $C_l^1\lambda_i^{l-1}\longrightarrow\infty$ $(l\to\infty)$，这是不可能的。反过来，若 $\lambda_i=1$ 且 $n_i=1$，则 $J_{n_i}(\lambda_i)=[1]$，故 $J_{n_i}^l(\lambda_i)\longrightarrow[1]$ $(l\longrightarrow\infty)$。若 $|\lambda_i|<1$，则由于 $\lim\limits_{l\to\infty}C_l^j\lambda_i^{l-j}=O$ 从而 $\lim\limits_{l\to\infty}[J_{n_i}(\lambda_i)]^l=O_{n_i}$。

为了强调上述引理的结论，故给出

推论4.2.6 如果Jordan 块 $J_{n_i}(\lambda_i)$ 幂收敛，且记

$$K=\lim_{l\to\infty}[J_{n_i}(\lambda_i)]^l$$

则有

$$K=\begin{cases}O_{n_i}, & |\lambda_i|<1\\ [1], & \lambda_i=1\end{cases}$$

有了上面的准备，便可证明下述的重要结论。

定理4.2.7 方阵 $A\in\mathbf{C}^{n\times n}$ 幂收敛，当且仅当 A 的所有特征值都满足：

1. $|\lambda|\leqslant 1$，
2. 若 $|\lambda|=1$，则 $\lambda=1$ 且相应的Jordan块是一阶的。

证明 据前面讨论，A 是幂收敛的，当且仅当它的 Jordan 标准形 J 中的每一个Jordan 块 $J_{n_i}(\lambda_i)$ $i=1,2,\cdots$，都是幂收敛的。再据引理4.2.5便得所需的结论。

与方阵幂收敛有关的一个概念是幂有界：对于给定的方阵 $A=(a_{ij})\in\mathbf{C}^{n\times n}$，若存在正数 M 使得对一切正整数 k，都有

$$|(A^k)_{ij}| \leqslant M, \qquad i, j = 1, 2, \cdots, n$$

则称A是幂有界的。很容易证明

定理4.2.8 若A是幂有界的，P为任一满秩阵，则$B = P^{-1}AP$也是幂有界的。

证明 因为对任意正整数k都有

$$B^k = P^{-1}A^kP, \qquad |(A^k)_{st}| \leqslant M$$

所以当$P = (p_{ij})$，$P^{-1} = (q_{ij})$，且

$$a = \max_{i,j} |p_{ij}|, \qquad b = \max_{i,j} |q_{ij}|$$

时，有

$$\begin{aligned} |(B^k)_{ij}| &= |\sum_s \sum_t q_{is}(A^k)_{st} p_{tj}| \\ &\leqslant \sum_s \sum_t |q_{is}||(A^k)_{st}||p_{tj}| \\ &\leqslant \max_{i,s} |q_{is}| M \max_{t,j} |p_{tj}| \sum_s \sum_t 1 \\ &= n^2 abM \triangleq K \end{aligned}$$

则因n，a，b，M均是不变的常数，故B是幂有界的。

这个定理表明，若$A \sim J$，则有

$$A\text{是幂有界的} \Longleftrightarrow J\text{是幂有界的。}$$

如果方阵A为块对角阵：

$$A = \operatorname{diag}(A_1, A_2, \cdots, A_k)$$

则由于

$$A^m = \operatorname{diag}(A_1^m, A_2^m, \cdots, A_k^m)$$

故容易看出

$$A\text{是幂有界的} \Longleftrightarrow \text{每个}A_i\text{是幂有界的。}$$

关于方阵的幂有界，有类似定理4.2.7的结论。

定理4.2.9 方阵$A \in \mathbf{C}^{n \times n}$是幂有界的，当且仅当$A$的特征值都满足

1. $|\lambda| \leqslant 1$,
2. 若$|\lambda| = 1$，则必有$\lambda = 1$且相应的Jordan块为一阶的。

证明 根据前面的讨论，只须证明 A 的每一个 Jordan 块 $J_{n_i}(\lambda_i)$幂有界的充分必要条件是：要么$|\lambda_i|<1$，要么$\lambda_i=1$且$n_i=1$。其证明与引理4.2.5的证明相仿。

这个定理表明方阵A的幂收敛与幂有界是完全等价的两个概念。

有了以上的准备，我们就可以研究矩阵级数了。

定义4.2.10 如果$\{A_k\}$是矩阵空间$\mathbf{C}^{m\times n}$中的一个无穷序列，则把

$$A_1+A_2+\cdots+A_k+\cdots$$

叫做由$\{A_k\}$生成的无穷级数，记为$\sum_{k=1}^{\infty}A_k$。对任一有限的m，则把$s_m=\sum_{k=1}^{m}A_k$叫做级数的部分和，而把$\{s_m\}$叫做级数$\sum_{k=1}^{\infty}A_k$的部分和序列。如果极限$\lim_{m\to\infty}s_m$存在，则称级数$\sum_{k=1}^{\infty}A_k$是收敛的。否则，便称之为发散的。如果对$\mathbf{C}^{m\times n}$上的某种范数$\|\cdot\|$，标量级数$\sum_{k=1}^{\infty}\|A_k\|$收敛，则称级数$\sum_{k=1}^{\infty}A_k$是绝对收敛的。

这里有与标量级数类似的结论。

定理4.2.11 若级数$\sum_{k=1}^{\infty}A_k$绝对收敛，则一定是收敛的。

证明 根据矩阵级数$\sum_{k=1}^{\infty}A_k$收敛的定义可知，$\sum_{k=1}^{\infty}A_k$是收敛的，等价于下面$m\cdot n$个标量级数

$$\sum_{k=1}^{\infty}(A_k)_{ij},\qquad 1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n$$

同时收敛。由于有限维空间上各种范数互相等价，不妨取矩阵的契比雪夫范数

$$\|A\|=\max_{i,j}|a_{ij}|$$

由于$\sum_{k=1}^{\infty}\|A_k\|$收敛，根据Cauchy准则有

$$\lim_{\substack{s,t\to\infty\\ s\geq t}}\sum_{k=t}^{s}\|A_k\|=0$$

所以对所有的$i=1, 2, \cdots, m$，$j=1, 2, \cdots, n$都有

$$\begin{aligned}\lim_{\substack{s,t\to\infty\\ s\geq t}}\left|\sum_{k=t}^{s}(A_k)_{ij}\right| &\leqslant \lim_{\substack{s,t\to\infty\\ s\geq t}}\sum_{k=t}^{s}|(A_k)_{ij}|\\ &\leqslant \lim_{\substack{s,t\to\infty\\ s\geq t}}\sum_{k=t}^{s}\max_{i,j}|(A_k)_{ij}|\\ &=\lim_{\substack{s,t\to\infty\\ s\geq t}}\sum_{k=t}^{s}\|A_k\|=0\end{aligned}$$

这表明每个标量级数都收敛，定理获证。

有时可以用上述定理来判定矩阵级数的收敛性。例如，对任给的方阵$A\in\mathbf{C}^{n\times n}$，级数

$$I+A+\frac{1}{2!}A^2+\cdots+\frac{1}{n!}A^n+\cdots+=\sum_{n=0}^{\infty}\frac{A^n}{n!}$$

是绝对收敛的，因为对任一种范数$\|\cdot\|$有

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{\|A^n\|}{n!}&=\|I\|+\|A\|+\frac{1}{2!}\|A^2\|+\cdots\\ &=e^{\|A\|}+\|I\|-1\end{aligned}$$

所以级数$\sum_{n=0}^{\infty}\frac{A^n}{n!}$也收敛，暂且记为$e^A=\sum_{n=0}^{\infty}\frac{A^n}{n!}$。

同样的方法可证级数

$$\sum_{n=0}^{\infty}\frac{(-1)^nA^{2n}}{(2n)!}\quad,\quad\sum_{n=0}^{\infty}\frac{(-1)^nA^{2n+1}}{(2n+1)!}$$

也是绝对收敛的，因而是收敛的。顺便再指出矩阵级数的几个常用性质：如果在矩阵空间$\mathbf{C}^{m\times n}$中，$\sum_{n=1}^{\infty}A_n=S$，$\sum_{n=1}^{\infty}B_n=T$，则有

1. $\sum_{n=1}^{\infty}(A_n+B_n)=S+T$;

2. 任取$\alpha\in\mathbf{C}$都有$\sum_{n=1}^{\infty}\alpha A_n=\alpha S$

3. 任取$x\in\mathbf{C}^n$都有$\sum_{n=1}^{\infty}A_n x=Sx$

这些性质的证明是很容易的。

第三节　幂級数及Lagrange-Sylvester定理

对给定的标量序列$\{\alpha_k\}$及方阵$A\in\mathbf{C}^{n\times n}$，所生成的矩阵级数$\sum_{k=0}^{\infty}\alpha_k A^k$叫做$A$的一个幂级数。它是数学中常用的一类级数。我们的目的是研究：在给定$\{\alpha_k\}$之后，方阵A满足什么条件才能使$\sum_{k=0}^{\infty}\alpha_k A^k$收敛?它又收敛到怎样的一个矩阵$S$?$S$如何计算?据定义，级数$\sum_{k=0}^{\infty}\alpha_k A^k$的收敛等价于$n^2$个标量级数$\sum_{k=0}^{\infty}\alpha_k(A^k)_{ij}$ $(i,j=1,2,\cdots,n)$的同时收敛。因此会想到矩阵级数$\sum_{k=0}^{\infty}\alpha_k A^k$的收敛与标量级数$\sum_{k=0}^{\infty}\alpha_k z^k$有一定的关系，我们试分析一下。

假定$A=PJP^{-1}$，其中

$$J=\mathrm{diag}(J_{n_1}(\lambda_1),\ J_{n_2}(\lambda_2),\ \cdots,\ J_{n_k}(\lambda_k))$$

据前面的分析有

$$\begin{aligned}A^l&=PJ^lP^{-1}\\&=P\mathrm{diag}(J_{n_1}^l(\lambda_1),J_{n_2}^l(\lambda_2),\ \cdots,\ J_{n_k}^l(\lambda_k))P^{-1}\end{aligned}$$

$$\begin{aligned}\sum_l\alpha_l A^l&=P(\sum_l\alpha_l J^l)P^{-1}\\&=P\mathrm{diag}(\sum_l\alpha_l J_{n_1}^l(\lambda_1),\ \sum_l\alpha_l J_{n_2}^l(\lambda_2),\ \cdots,\\&\qquad\sum_l\alpha_l J_{n_k}^l(\lambda_k))P^{-1}\end{aligned}$$

因为当$l \geqslant n_i$时，有

$$J_{n_i}^l(\lambda_i)=(\lambda_i I_{n_i}+U_{n_i})^l$$

$$=\sum_{j=0}^{n_i-1} C_l^j \lambda_i^{l-j} U_{n_i}^j$$

$$=\begin{pmatrix} \lambda_i^l & C_l^1\lambda_i^{l-1} & \cdots\cdots & C_l^{l-n_i+1}\lambda_i^{n_i-1} \\ & \lambda_i^l & C_l^1\lambda_i^{l-1} & \vdots \\ & & \lambda_i^l & \vdots \\ & & \ddots & \vdots \\ & & & C_l^1\lambda_i^{l-1} \\ & & & \lambda_i^l \end{pmatrix}$$

故若A的特征值 λ_i 落在标量级 数 $f(z)=\sum_l \alpha_l z^l$ 的收斂圆内，则便有 $f(\lambda_i)=\sum_l \alpha_l \lambda_i^l, \frac{1}{j!}f^{(j)}(\lambda_i)=\sum_l \alpha_l C_l^j \lambda_i^{l-j}$，从而便得到

$$\sum_l \alpha_l J_{n_i}^l(\lambda_i)=\sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_i)}{j!} U_{n_i}^j$$

$$=\begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots\cdots\cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \vdots \\ & & \ddots & \vdots \\ & & & f'(\lambda_i) \\ & & & f(\lambda_i) \end{pmatrix}$$

由此可得

$$\sum \alpha_l A^l = P\left[\bigoplus_{i=1}^{k} \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_i)}{j!} U_{n_i}^{j}\right] P^{-1}$$

其中 $\bigoplus_{i=1}^{k} B_i = \operatorname{diag}(B_1, B_2, \cdots, B_k)$

例如，当 $A = P\operatorname{diag}(J_4(3), J_3(1))P^{-1}$ 时，则有

$$\sum \alpha_l A^l = P\begin{bmatrix} f(3) & f'(3) & \frac{1}{2!}f''(3) & \frac{1}{3!}f'''(3) & & & \\ & f(3) & f'(3) & \frac{1}{2!}f''(3) & & & \\ & & f(3) & f'(3) & & & \\ & & & f(3) & & & \\ & & & & f(1) & f'(1) & \frac{1}{2!}f''(1) \\ & & & & & f(1) & f'(1) \\ & & & & & & f(1) \end{bmatrix} P^{-1}$$

特别是当 $\sum_l^{\infty} \alpha_l A^l = \sum_l^{\infty}\left(\frac{A}{6}\right)^l$ 时， $f(z) = \sum_l^{\infty}\left(\frac{z}{6}\right)^l = \left(1 - \frac{z}{6}\right)^{-1}$ 则 $f'(z) = \frac{1}{6}\left(1 - \frac{z}{6}\right)^{-2}$， $f''(z) = \frac{1}{18}\left(1 - \frac{z}{6}\right)^{-3}$， $f'''(z) = \frac{1}{36}\left(1 - \frac{z}{6}\right)^{-4}$ 故而又有

$$\sum_l^{\infty} \alpha_l A^l = P\begin{bmatrix} 2 & 2/3 & 2/9 & 2/27 & & & \\ & 2 & 2/3 & 2/9 & & & \\ & & 2 & 2/3 & & & \\ & & & 2 & & & \\ & & & & 6/5 & 6/25 & 6/125 \\ & & & & & 6/5 & 6/25 \\ & & & & & & 6/5 \end{bmatrix} P^{-1}$$

总结上面的分析，便可得出结论：

定理4.3.1 （Lagrange-Sylvester）假定

1. R是标量级数$\sum_l \alpha_l z^l$的收敛半径，且对所有的$|z|<R$，有$f(z)=\sum_l \alpha_l z^l$；

2. 方阵$A\in \mathbf{C}^{n\times n}$的所有特征值$\lambda$都满足$|\lambda|<R$；

3. $A=PJP^{-1}$，$J=\bigoplus_{i=1}^{k} J_{n_i}(\lambda_i)$为$A$的 Jordan 标准形，$U_{n_i}=J_{n_i}(0)$，$f^{(j)}$为$f(z)$的$j$阶导数。

则有

$$\sum_{l=0}^{\infty}\alpha_l A^l=P\left[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{f^{(j)}(\lambda_i)}{j!}U_{n_i}^{j}\right]P^{-1}$$

这便是著名的Lagrange-Sylvester定理。

推论4.3.2 若方阵A的所有特征值λ都落在标量级数$\sum\alpha_l z^l$的收敛圆之内，或者A的谱半径$\rho(A)$（$=\max\limits_{1\leqslant i\leqslant n}|\lambda_i|$）小于级数$\sum\alpha_l z^l$的收敛半径$R$，则矩阵$A$的幂级数$\sum_{l=0}^{\infty}\alpha_l A^l$是收敛的。特别当

$$f(z)=\sum_l \alpha_l z^l \qquad |z|<R$$

时，我们也形式地把矩阵幂级数记为

$$f(A)=\sum_l \alpha_l A^l$$

推论4.3.3 若方阵A的特征值λ有一个落在标量级数$\sum_l \alpha_l z^l$的收敛圆之外，或A的谱半径$\rho(A)$大于$\sum_l \alpha_l z^l$的收敛半径，则幂级数$\sum_l \alpha^l A^l$是发散的。

推论4.3.4 若标量级数$\sum_{l=0}^{\infty}\alpha_l(z-\lambda_0)^l$的收敛半径为$R$，方阵$A\in \mathbf{C}^{n\times n}$的所有特征值都满足

$$|\lambda_i-\lambda_0|<R, \qquad i=1, 2, \cdots, n$$

则幂级数$\sum\alpha_l(A-\lambda_0 I)^l$是收敛的。

证明 令$B=A-\lambda I$，则因为

$$\rho(B)=\max_{1\leq i\leq n}|\lambda_i-\lambda_0|<R$$

据推论4.3.2知$\sum_l \alpha_l B^l$收敛。

从上面的定理和推论可以看出：若标量级数$\sum_l \alpha_l z^l$在整个复平面上收敛，则对任意的方阵$A\in\mathbf{C}^{n\times n}$，幂级数$\sum_l \alpha_l A^l$都是收敛的。据此可知，对任一$A\in\mathbf{C}^{n\times n}$，下列级数

$$\sum_{n=0}^{\infty}\frac{A^n}{n!},\qquad \sum_{n=0}^{\infty}(-1)^n\frac{A^{2n+1}}{(2n+1)!},\qquad \sum_{n=0}^{\infty}(-1)^n\frac{A^{2n}}{(2n)!}$$

都是收敛的。

我们知道，如果标量级数$\sum_l \alpha_l z^l$的收敛半径为R，则对任意的$|z|<R$，$\sum_l \alpha_l z^l$都是收敛的，所以便定义了一个函数

$$f(z)=\sum_l \alpha_l z^l,\qquad |z|<R$$

同样，当我们取$\mathbf{C}^{n\times n}$的子集

$$D(f)\triangleq\{A\in\mathbf{C}^{n\times n}\,|\,\rho(A)<R\}$$

则对每个$B\in D(f)$，矩阵的幂级数$\sum_l \alpha_l B^l$也是收敛的。这样

$$f(A)=\sum_l \alpha_l A^l\qquad A\in D(f)$$

便在$D(f)$上定义了一个以矩阵A为变量而取值也为矩阵的函数，换句话说，f是一个映射：

$$f:\ D(f)\longrightarrow\mathbf{C}^{n\times n}$$

$$A\longmapsto f(A)=\sum_l \alpha_l A^l$$

一般来说，把$f(A)$叫做标量函数的矩阵推广或矩阵函数，如

$$e^A=\sum_{n=0}^{\infty}\frac{A^n}{n!},\qquad A\in\mathbf{C}^{n\times n}$$

$$\sin A=\sum_{n=0}^{\infty}(-1)^n\frac{A^{2n+1}}{(2n+1)!},\qquad A\in\mathbf{C}^{n\times n}$$

$$\cos A=\sum_{n=0}^{\infty}(-1)^n\frac{A^{2n}}{(2n)!},\qquad A\in\mathbf{C}^{n\times n}$$

$$(I-A)^{-1}=\sum_{n=0}^{\infty}A^n, \qquad \rho(A)p<1$$

$$(I+A)^{-1}=\sum_{n=0}^{\infty}(-A)^n, \qquad \rho(A)<1$$

仿照标量级数的公式，还可以定义其它的矩阵函数。

关于三角函数、指数函数，对数函数有一系列的恒等式和关系式，例如

$$\sin^2 z+\cos^2 z=1, \qquad \sin 2z=2\sin z\cos z$$

自然会问：对于任一 $A\in\mathbf{C}^{n\times n}$，是否也有

$$\sin^2 A+\cos^2 A=I, \qquad \sin 2A=2\sin A\cos A$$

这个问题的回答是肯定的。虽说据定义可直接证明，但由于级数的相乘极其复杂，直接验证会有技术上的困难。现在寻求一种间接的方法，它不仅可以解决上述问题，而且在证明矩阵函数的一般性质时也颇为有用。为此，先证明一个引理

引理4.3.5 若对一切$|z|<R$有

$$f(z)=\sum_{j=0}^{\infty}\alpha_j z^j, \qquad g(z)=\sum_{j=0}^{\infty}\beta_j z^j,$$

$$s(z)=f(z)+g(z) \qquad p(z)=f(z)g(z)$$

则对一切满足$\rho(A)<R$的$A\in\mathbf{C}^{n\times n}$，也有

$$s(A)=f(A)+g(A), \qquad p(A)=f(A)g(A)$$

证明　因为对一切$|z|<R$有

$$s(z)=\sum_{j=0}^{\infty}(\alpha_j+\beta_j)z^j$$

故据定理4.3.1的推论4.3.2有

$$s(A)=\sum_{j=0}^{\infty}(\alpha_j+\beta_j)A^j, \qquad A\in\mathbf{C}^{n\times n}\text{且}\rho(A)<R$$

从而据定理4.2.4有

$$s(A)=\sum_{j=0}^{\infty}\alpha_j A^j+\sum_{j=0}^{\infty}\beta_j A^j=f(A)+g(A)$$

至于$p(A)=f(A)g(A)$，据定理4.3.1有

$$f(A)g(A)=P\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{f^{(j)}(\lambda_i)}{j!}U_{n_i}^{j}\Big]$$

$$\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{g^{(j)}(\lambda_i)}{j!}U_{n_i}^{j}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\sum_{r=0}^{n_i-1}\frac{f^{(j)}(\lambda_i)g^{(r)}(\lambda_i)}{j!r!}$$

$$U_{n_i}^{j+r}\Big]P^{-1}$$

若令$l=j+r$，并注意到当$l\geqslant n_i$时$U_{n_i}^{l}=O$，便有

$$f(A)g(A)=P\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\sum_{l=j}^{n_i-1}$$

$$\frac{f^{(j)}(\lambda_i)g^{(l-j)}(\lambda_i)}{j!(l-j)!}U_{n_i}^{l}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{l=0}^{n_i-1}\sum_{j=0}^{l}$$

$$\frac{f^{(j)}(\lambda_i)g^{(l-j)}(\lambda_i)}{j!(l-j)!}U_{n_i}^{l}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{l=0}^{n_i-1}\sum_{j=0}^{l}C_l^j$$

$$\frac{f^{(j)}(\lambda_i)g^{(l-j)}(\lambda_i)}{l!}U_{n_i}^{l}\Big]P^{-1}$$

这一结果与标量级数相乘的公式

$$f(z)g(z)=\sum_{l=0}^{\infty}\sum_{j=0}^{l}\alpha_j\beta_{l-j}z^l$$

极为相似。另一方面，据Lagrange-Sylvester定理

$$p(A)=P\Big[\bigoplus_{i=1}^{k}\sum_{l=0}^{n_i-1}\frac{p^{(l)}(\lambda_i)}{l!}U_{n_i}^{l}\Big]P^{-1}$$

再据函数乘积求导的Leibniz法则有

$$p^{(l)}(z)=(f(z)g(z))^{(l)}=\sum_{j=0}^{l}C_l^j f^{(j)}(z)g^{(l-j)}(z)$$

于是有

$$p(A)=P\left[\bigoplus_{i=1}^{k}\sum_{l=0}^{n_i-1}\frac{p^{(l)}(\lambda_i)}{l!}U_{n_i}^l\right]P^{-1}$$

$$=P\left[\bigoplus_{i=1}^{k}\sum_{l=0}^{n_i-1}\sum_{j=0}^{l}C_l^j\frac{f^{(j)}(\lambda_i)g^{(l-j)}(\lambda_i)}{l!}U_{n_i}^l\right]P^{-1}$$

与前面$f(A)g(A)$的表达式比较，从而证明了

$$p(A)=f(A)g(A)\ 。$$

为了说明引理所述的方法，现在来证明

$$\sin^2A+\cos^2A=I,\qquad 2\sin A\cos A=\sin 2A$$

设 $\sin^2z$，$\cos^2z$的幂级数展开式为

$$\sin^2z=\sum\alpha_jz^j,\qquad \cos^2z=\sum\beta_jz^j$$

据引理有

$$\sin^2A=\sum\alpha_jA^j,\qquad \cos^2A=\sum\beta_jA^j$$

从而有

$$\sin^2A+\cos^2A=\sum(\alpha_j+\beta_j)A^j \tag{*}$$

但由于恒等式

$$\sin^2z+\cos^2z=\sum(\alpha_j+\beta_j)z^j=1$$

故比较 z 的同次幂的系数，应有

$$\alpha_0+\beta_0=1,\qquad \alpha_j+\beta_j=0,\qquad j=1,\ 2,\ \cdots$$

从而式（*）的右端级数只有一项I，这便证明了

$$\sin^2A+\cos^2A=I$$

类似地设函数 $\sin z\cos z$ 的幂级数展开式为

$$\sin z\cos z=\sum\delta_jz^j$$

从而有

$$\sin A\cos A=\sum\delta_jA^j$$

又由于$\sin 2z=2\sin z\cos z=\sum 2\delta_jz^j$，故有

$$\sin 2A=\sum 2\delta_j A^j=2\sum \delta_j A^j$$

对$\sin A\cos A$及$\sin 2A$的表达式加以比较便有

$$\sin 2A=2\sin A\cos A$$

引理4.3.5是关于两个标量函数的和与积的矩阵推广，其实据归纳法可证：

推论4.3.6 设

$$f_i(z)=\sum_j a_j^{(i)} z^j, \quad i=1, 2, \cdots, m, \quad |z|<R$$

$$f_i(A)=\sum_j a_j^{(i)} A^j, \quad A\in \mathbf{C}^{n\times n}\text{且}\rho(A)<R$$

则

1. 若$\prod_{i=1}^{m} f_i(z)$的幂级数展开式为$\sum_j \beta_j z^j$，便有

$$\prod_{i=1}^{m} f_i(A)=\sum \beta_j A^j, \quad \rho(A)<R$$

2. $\sum_{i=1}^{m} \delta_i f_i(z)$的矩阵推广为$\sum_{i=1}^{m} \delta_i f_i(A)$。

有了上述引理及推论，可以证明一个比较有用的定理。

定理4.3.7 设f_1，f_2，…，f_m为一元标量函数且

$$f_i(z)=\sum_{j=0}^{\infty} a_j^{(i)} z^j, \qquad |z|<R, \quad i=1, 2, \cdots, m$$

$\varphi(x_1, x_2, \cdots, x_m)$为$m$个变元$x_1$，$x_2$，…，$x_m$的多项式。如果对一切$|z|<R$都有

$$\varphi(f_1(z), f_2(z), \cdots, f_m(z))=0$$

则对满足$\rho(A)<R$的一切$A\in \mathbf{C}^{n\times n}$也有

$$\varphi(f_1(A), f_2(A), \cdots, f_m(A))=0。$$

证明 令$\psi(z)=\varphi(f_1(z),f_2(z),\cdots,f_m(z))$，则$\psi(z)$为诸$f_i(z)$的幂积的线性组合，据推论4.3.6，$\psi(z)$有矩阵推广$\psi(A)=\varphi(f_1(A), f_2(A), \cdots, f_m(A))$，并且当$|z|<R$及$\rho(A)<R$时有

$$\psi(z)=\sum_{j=0}^{\infty} \beta_j z^j, \qquad \psi(A)=\sum_{j=0}^{\infty} \beta_j A^j$$

因为 $\psi(z)=0$，故 $\beta_j=0$，$j=0, 1, \cdots$，从而也有 $\psi(A)=\sum_{j=0}^{\infty} 0\, A^j=0$。

利用上述定理，可以很方便地证明某些矩阵函数的等式。例如，在定理4.3.7中取 $f_1=\sin z$，$f_2=\cos z$，$f_3=\sin 2z$，$\varphi_1(x_1, x_2)=x_1^2+x_2^2-1$，$\varphi_2(x_1, x_2, x_3)=2x_1x_2-x_3$，便立即得到

$$\sin^2 A+\cos^2 A-I=0, \qquad 2\sin A\cos A-\sin 2A=0$$

如果在定理4.3.7中取 $\varphi(x_1, x_2)=1-x_1x_2$，$f_1(z)=1-z$，$f_2(z)=(1-z)^{-1}$，并且当 $|z|<1$ 时有

$$f_2(z)=(1-z)^{-1}=\sum_{j=0}^{\infty} z^j$$

$$\varphi(f_1(z), f_2(z))=0$$

故据定理4.3.7的结论，应有

$$0=\varphi(f_1(A), f_2(A))=I-(I-A)\sum_{j=0}^{\infty} A^j \quad , \rho(A)<1$$

从而可知，对满足 $\rho(A)<1$ 的一切 $A\in \mathbf{C}^{n\times n}$ 有

$$(I-A)^{-1}=\sum_{j=0}^{\infty} A^j$$

同样的方法，很容易证明对一切 $A\in \mathbf{C}^{n\times n}$ 有

$$\cos 2A=\cos^2 A-\sin^2 A, \qquad \cos^2 A=\frac{1}{2}(I+\cos 2A)$$

$$\sin^2 A=\frac{1}{2}(I-\cos 2A)。$$

以上我们是通过标量函数的幂级数展开

$$f(z)=\sum_j \alpha_j z^j, \qquad |z|<R$$

从形式上来定义矩阵函数

$$f(A)=\sum_j \alpha_j A^j, \qquad \rho(A)<R$$

至于这种函数有哪些性质，给定了 A 后如何计算 $f(A)$，以及矩阵函数有什么应用，都是要进一步考虑的问题。不过，Lagrange-

Sylvester定理倒是提供了计算$f(A)$的一条途径，但是首先要知道A的Jordan 标准形及相似变换阵 P。这需要大量的准备工作，有时甚至是不可能做到的。当然，对简单的或低阶的A，还是可行的。例如，知道了

$$A=\begin{bmatrix}1 & 4\\ 3 & 2\end{bmatrix}=P\begin{bmatrix}5 & 0\\ 0 & -2\end{bmatrix}P^{-1}, \quad P=\begin{bmatrix}1 & -4\\ 1 & 3\end{bmatrix}$$

则据 Lagrange-Syevester 定理有

$$e^A=P\begin{bmatrix}e^5 & 0\\ 0 & e^{-2}\end{bmatrix}P^{-1}=\frac{1}{7}\begin{bmatrix}3e^5+4e^{-2} & 4e^5-4e^{-2}\\ 3e^5-3e^{-2} & 4e^5+3e^{-2}\end{bmatrix}$$

为了避免计算J及P,现在另觅途径。首先让我们分析一下，如何对$f(A)=\sum\limits_{i=0}^{\infty}a_iA^i$加以变形，以达到求出$f(A)$的目的。我们从 Cayley-Hamilton 定理知道，每个矩阵$A\in\mathbf{C}^{n\times n}$的$n$次或高于$n$次的幂$A^k$ $(k\geqslant n)$ 都能表示为A的一个次数不超过$n-1$的多项式。既然矩阵函数 $f(A)=\sum a_iA^i$ 的级数表达式收敛，而 A^k $(k\geqslant n)$ 又可表为A的多项式，则可望 $f(A)$ 具有某一多项式的形式，不妨设$f(A)=p(A)$，而

$$p(\lambda)=\beta_0+\beta_1\lambda+\cdots+\beta_l\lambda^l$$

这样问题就转化为确定多项式的系数β_i $(i=0,1,\cdots,l)$。这里还有$p(\lambda)$的次数l的确定问题。其实，Cayley-Hamilton定理只是说A的特征多项式是A的一个零化多项式。如果A有比特征多项式次数低的零化多项式，则可望 $p(\lambda)$ 的次数更低，因此以A的最小多项式的次数来界定$p(\lambda)$的次数。设A的最小多项式为

$$m_A(\lambda)=\prod_{i=1}^{q}(\lambda-\lambda_i)^{l_i}$$

其中λ_1，λ_2，…，λ_q为A的特征值且互异。这时可设 $p(\lambda)$次数 l 低于$m_A(\lambda)$的次数一次。据Lagrange-Sylvester定理应有

$$f(A)=P\left[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{f^{(j)}(\lambda_i)}{j!}U_{n_i}^{j}\right]P^{-1}$$

$$p(A)=P\left[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{p^{(j)}(\lambda_i)}{j!}U_{n_i}^{j}\right]P^{-1}$$

但据假设$f(A)=p(A)$，故应有

$$p^{(j)}(\lambda_i)=f^{(j)}(\lambda_i),\ j=0,\ 1,\cdots,n_i-1,\ i=1,\ 2,\ \cdots,k。$$

因此，只要求出λ_i及$f^{(j)}(\lambda_i)$代入到

$$p^{(j)}(\lambda)=[\beta_0+\beta_1\lambda+\cdots+\beta_l\lambda^l]^{(j)}$$

两端，便可定出β_0，β_1，…，β_l，从而得到

$$f(A)=\beta_0I+\beta_1A+\cdots+\beta_lA^l$$

再以

$$A=\begin{bmatrix}1 & 4\\ 3 & 2\end{bmatrix}$$

为例，计算e^A。由于$m_A(\lambda)=(\lambda-5)(\lambda+2)$，故设$p(\lambda)=\beta_0+\beta_1\lambda$，令它满足

$$p(5)=e^5,\qquad p(-2)=e^{-2}$$

从而解方程组

$$\begin{cases}\beta_0+5\beta_1=e^5\\ \beta_0-2\beta_1=e^{-2}\end{cases}$$

可得 $\beta_0=\frac{1}{7}(2e^5+5e^{-2})$， $\beta_1=\frac{1}{7}(e^5-e^{-2})$

最后便求得

$$e^A=\frac{1}{7}(2e^5+5e^{-2})I+\frac{1}{7}(e^5-e^{-2})A$$

最后以

$$A=\begin{bmatrix}0 & -1\\ 4 & 4\end{bmatrix}$$

为例，计算$\sin^{-1}\left(\frac{A}{4}\right)$。很容易求出$m_A(\lambda)=(\lambda-2)^2$。现设$p(\lambda)=\beta_0+\beta_1\lambda$，并满足

$$p(2)=\beta_0+2\beta_1=\sin^{-1}\frac{1}{2}=\frac{\pi}{6}$$

$$p'(2)=\beta_1=\frac{1}{4}\left[1-\left(\frac{1}{2}\right)^2\right]^{-1/2}=\frac{\sqrt{3}}{6}$$

从而得到 $p(\lambda)=\frac{1}{6}[(\pi-2\sqrt{3})+\sqrt{3}\lambda]$。最后便求得

$$\sin^{-1}\left(\frac{A}{4}\right)=p(A)$$

$$=\frac{1}{6}\begin{bmatrix}\pi-2\sqrt{3} & -\sqrt{3}\\ 4\sqrt{3} & \pi+2\sqrt{3}\end{bmatrix}$$

现在再看一个阶数稍高的矩阵

$$A=\begin{pmatrix}-2 & 2 & -2 & 4\\ -1 & 2 & -1 & 1\\ 0 & 0 & 1 & 0\\ -2 & 1 & -1 & 4\end{pmatrix}$$

要求$\cos\pi A$。计算可知A的最小多项式为

$$m_A(\lambda)=\lambda^3-4\lambda^2+5\lambda-2=(\lambda-1)^2(\lambda-2)$$

令$p(\lambda)=\beta_0+\beta_1\lambda+\beta_2\lambda^2$，且满足

$$p(2)=\beta_0+2\beta_1+4\beta_2=\cos2\pi=1$$

$$p(1)=\beta_0+\beta_1+\beta_2=\cos\pi=-1$$

$$p'(1)=\beta_1+2\beta_2=-\pi\sin\pi=0$$

解方程组可得 $\beta_0=1$，$\beta_1=-4$，$\beta_2=2$。于是便求得

$$\cos\pi A=p(A)=I-4A+2A^2=\begin{pmatrix}-3 & 0 & 0 & 4\\ 0 & -1 & 0 & 0\\ 0 & 0 & -1 & 0\\ -2 & 0 & 0 & 3\end{pmatrix}$$

从上述的分析与计算可以看出，这种方法避开了对J及P的计算，而去求$p(A)$。而欲求 $p(\lambda)$，只须计算A的最小多项式及数

值 $f^{(j)}(\lambda_i)$ $(i=1, 2, \cdots, q, j=0, 1, \cdots, l_i-1)$。我们把上面的分析综合为下面的定理：

定理4.3.8 设方阵$A\in\mathbf{C}^{n\times n}$的最小多项式为

$$m_A(\lambda)=\prod_{i=1}^{q}(\lambda-\lambda_i)^{l_i}$$

且 $s=\sum_{i=1}^{q}l_i$。而 $f(A)=\sum_{i=0}^{\infty}\alpha_i A^i$ 为 A的收敛级数。则必存在一个次数低于 s 的多项式$p(\lambda)$，它与$f(\lambda)$在A的谱点λ_1，λ_2，$\cdots$，λ_q上是一致的，即

$$p^{(j)}(\lambda_i)=f^{(j)}(\lambda_i),\quad 0\leqslant j<l_i,\quad i=1, 2, \cdots, q$$

并且使得

$$f(A)=p(A)$$

上述定理中的 $p(\lambda)$ 又叫做 $f(\lambda)$ 的在A的谱上一致的多项式，简称谱上一致多项式。

推论4.3.9 若 $A\in\mathbf{C}^{n\times n}$ 为块对角阵

$$A=\mathrm{diag}(A_1, A_2, \cdots, A_t)$$

且级数 $f(A)=\sum_{i=0}^{\infty}\alpha_i A^i$收敛，则有

$$f(A)=\mathrm{diag}(f(A_1), f(A_2), \cdots, f(A_t))$$

证明 显然有

$$\begin{aligned}f(A)&=\lim_{k\to\infty}\sum_{i=0}^{k}\alpha_i A^i\\&=\lim_{k\to\infty}\sum_{i=0}^{k}\mathrm{diag}(\alpha_i A_1^i, \alpha_i A_2^i, \cdots, \alpha_i A_t^i)\\&=\mathrm{diag}\left(\lim_{k\to\infty}\sum_{i=0}^{k}\alpha_i A_1^i, \cdots, \lim_{k\to\infty}\sum_{i=0}^{k}\alpha_i A_t^i\right)\\&=\mathrm{diag}(f(A_1), \cdots, f(A_t))\end{aligned}$$

推论4.3.10 若

$$A=J=\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}\in \mathbf{C}^{n\times n}$$

且 $f(z)=\sum_{i=0}^{\infty}\alpha_i A^i$的收敛半径为$R$，又$|\lambda|<R$，则据Lagrange-Sylvester定理有

$$f(A)=f(J)=\sum_{i=0}^{\infty}\alpha_i J^i$$

$$=\begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \frac{1}{2!}f''(\lambda) \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}$$

从上述推论可知，当A相似于对角阵时：

$$A=P\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)P^{-1}$$

而级数 $f(z)=\sum_{i=0}^{\infty}\alpha_i z^i$的收敛半径为$R$，且$\rho(A)<R$，则立即

可得

$$f(A)=P\mathrm{diag}(f(\lambda_1),\ f(\lambda_2),\ \cdots,\ f(\lambda_n))P^{-1}$$

特别有

$$e^{A}=P\mathrm{diag}(e^{\lambda_1},\ e^{\lambda_2},\ \cdots,\ e^{\lambda_n})P^{-1}$$

$$e^{tA}=P\mathrm{diag}(e^{\lambda_1 t},\ e^{\lambda_2 t},\cdots,\ e^{\lambda_n t})P^{-1}$$

$$\sin A=P\mathrm{diag}\ (\sin\lambda_1,\ \sin\lambda_2,\ \cdots,\ \sin\lambda_n)P^{-1}$$

$$\sin(tA)=P\mathrm{diag}\ (\sin\lambda_1 t,\ \sin\lambda_2 t,\ \cdots,\ \sin\lambda_n t)P^{-1}$$

$$\cos A=P\mathrm{diag}\ (\cos\lambda_1,\ \cos\lambda_2,\ \cdots,\ \cos\lambda_n)P^{-1}$$

$$\cos(tA)=P\mathrm{diag}\ (\cos\lambda_1 t,\ \cos\lambda_2 t,\ \cdots,\ \cos\lambda_n t)P^{-1}$$

利用谱上一致多项式计算矩阵函数是计算矩阵函数的一种方法，但决不是唯一的；用最小多项式来确定谱上一致多项式也不是最有效的方法。因为给定A，求$m_A(\lambda)$也不是轻而易举的。所以在数值计算中经常采用逐次逼近的方法。尽管如此，Lagrange-Sylvester定理给出的结果在理论上及工程应用上都仍是重要的。

第四节　函数矩阵的微分与积分，矩阵的标量函数的微分

在研究微分方程组时，为了简化为对矩阵的研究，需要考虑以函数为元素的矩阵（函数矩阵）的微分与积分。在研究某些优化问题时，则要碰到某些矩阵标量函数的求导。这里作简单的介绍。

如果矩阵 $A=(a_{ij})$ 的每个元素都是变量 t 的函数$a_{ij}(t)$，则称A为一个函数矩阵，记之为$A(t)$。若 $t\in(a,\ b)$，则称 $A(t)$ 是定义在$(a,\ b)$上的；又若每个$a_{ij}(t)$在$(a,\ b)$上连续、可微、可积，则称 $A(t)$在$(a,\ b)$上是连续、可微、可积。例如，当 $t\in[1,\ 2]$时，

$$A(t)=\begin{pmatrix}\sin t & \cos t & t\\ \dfrac{\sin t}{t} & e^{t} & t^2\\ 1 & 0 & t^3\end{pmatrix}\quad x(t)=\begin{pmatrix}e^{t}\\ \dfrac{1}{t^2}\\ 1\end{pmatrix}\quad y(t)=\begin{pmatrix}\sin\dfrac{\pi}{t}\\ \sin\pi t\end{pmatrix}$$

都是可微、可积的函数矩阵。严格定义如下：

定义4.4.1 若对一切 i，j 都有

$$\lim_{t\to a} a_{ij}(t)=a_{ij}<\infty$$

则称$A(t)$当$t\to a$时以$A=(a_{ij})$为极限，并记之为

$$\lim_{t\to a}A(t)=A$$

其中 a 可以是$\pm\infty$。

定义4.4.2 如果$A(t)$在$t=a$处有定义且

$$\lim_{t\to a}A(t)=A(a)$$

则称$A(t)$在$t=a$处连续。

定义4.4.3 如果极限

$$\lim_{\Delta t\to 0}(A(t+\Delta t)-A(t))/\Delta t$$

存在，则称$A(t)$在t处可微，并把上述极限叫做 $A(t)$ 的导数，记之为$A'(t)$或 $\dot{A}(t)$，$\frac{dA(t)}{dt}$。

据定义可知有

$$A'(t)=(a'_{ij}(t))$$

并且有如下结论：

定理4.4.4

1． 若$A(t)$，$B(t)$为同阶可微矩阵，则

$$\frac{d}{dt}(A(t)+B(t))=\frac{d}{dt}A(t)+\frac{d}{dt}B(t)$$

2． 若$A(t)$，$B(t)$分别为$m\times n$，$n\times l$ 阶可微矩阵，则

$$\frac{d}{dt}(A(t)B(t))=\left(\frac{d}{dt}A(t)\right)B(t)+A(t)\frac{d}{dt}B(t)$$

3． 若$A(u)$可微且$u=f(t)$关于 t 可微，则有

$$\frac{d}{dt}A(f(t))=f'(t)\frac{d}{du}A(u)$$

4． 若$A(t)$与$A^{-1}(t)$皆可微，则

$$\frac{d}{dt}(A^{-1}(t)) = -A^{-1}(t)\left(\frac{d}{dt}A(t)\right)A^{-1}(t)$$

证明　1，2，3是明显的，只证4。由于

$$A(t)A^{-1}(t)=I$$

上式两端关于 t 同时求导有

$$A'(t)A^{-1}(t)+A(t)\frac{d}{dt}(A^{-1}(t))=0$$

从而得到

$$\frac{d}{dt}(A^{-1}(t)) = -A^{-1}(t)\left(\frac{d}{dt}A(t)\right)A^{-1}(t)$$

注意：由于矩阵乘法一般是不能交换次序的，而函数矩阵 $A'(t)$ 与 $A^{-1}(t)$ 一般也是不可换的，所以4中的公式与标量函数的求导有差别。

类似地可定义$A(t)$的积分如下：

定义4.4.5　若$A(t)$的每个元素$a_{ij}(t)$都在$[a, b]$上可积，则称$A(t)$在$[a, b]$上是可积的，并规定

$$\int_a^b A(t)dt = \left[\int_a^b a_{ij}(t)\,dt\right]$$

据定义及微分积分的知识易知：

定理4.4.6

1．若$A(t)$在$[a, b]$上连续，则对任一$t\in(a,b)$，$\int_a^t A(\tau)d\tau$ 可微，且有

$$\frac{d}{dt}\left(\int_a^t A(\tau)d\tau\right)=A(t)$$

2．若$A(t)$在$[a, b]$上连续可微，则有

$$\int_a^t\left(\frac{d}{ds}A(s)\right)ds = A(t)-A(a)\qquad t\in[a, b]$$

上面是把标量函数的某些概念形式上推广到函数矩阵，没有什么新的内容，只是记法上的说明。这些内容下一节要用。现在

介绍某些矩阵标量函数的微分，并给出应用的例子。在众多的矩阵标量函数中，以矩阵的迹、行列式、范数最为重要，因为它们有很多应用，且有明确的表达式。

定理4.4.7 若 $A\in\mathbf{R}^{n\times m}$ 是给定的矩阵，$X\in\mathbf{R}^{m\times n}$ 是矩阵变量，X的标量函数为

$$\varphi(X)=\mathrm{tr}(AX)$$

则有

$$\frac{\partial\varphi}{\partial X}\triangleq\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=A^T.$$

证明 因为据迹的定义有

$$\varphi(X)=\mathrm{tr}(AX)=\sum_{s=1}^{n}(AX)_{ss}$$

$$=\sum_{s=1}^{n}\sum_{k=1}^{m}a_{sk}x_{ks}$$

故有

$$\frac{\partial\varphi}{\partial x_{ij}}=a_{ji},\qquad 1\leqslant i\leqslant m,\qquad 1\leqslant j\leqslant n$$

从而得到

$$\frac{\partial\varphi}{\partial X}\triangleq\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=[a_{ji}]=A^T$$

定理4.4.8 若 $B\in\mathbf{R}^{m\times n}$ 是给定的矩阵，$X\in\mathbf{R}^{m\times n}$ 是矩阵变量，矩阵的标量函数为

$$\varphi(X)=\mathrm{tr}(X^TBX)$$

则有

$$\frac{\partial\varphi}{\partial X}\triangleq\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=BX+B^TX$$

特别当B为对称阵时，$\dfrac{\partial\varphi}{\partial X}=2BX$。

证明 由于

$$(X^T BX)_{ss}=\sum_{l=1}^{m}\sum_{k=1}^{m}x_{ks}b_{kl}x_{ls}$$

$$\varphi(X)=\mathrm{tr}(X^T BX)=\sum_{s=1}^{n}\sum_{l=1}^{m}\sum_{k=1}^{m}x_{ks}b_{kl}x_{ls}$$

故有

$$\frac{\partial\varphi}{\partial x_{ij}}=\sum_{l=1}^{m}b_{il}x_{lj}+\sum_{k=1}^{m}b_{ki}x_{kj}$$

$$\frac{\partial\varphi}{\partial X}=\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=BX+B^T X$$

定理4.4.9 若$A\in\mathbf{R}^{n\times m}$, $B\in\mathbf{R}^{n\times n}$是给定的矩阵，$X\in\mathbf{R}^{m\times n}$ 是矩阵变量，标量函数为

$$\varphi(X)=\mathrm{tr}(AXB)$$

则有

$$\frac{\partial\varphi}{\partial X}\triangleq\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=A^T B^T$$

证明 据迹的定义写出

$$\varphi(X)=\sum_{s=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}a_{sk}x_{kl}b_{ls}$$

从而知

$$\frac{\partial\varphi}{\partial x_{ij}}=\sum_{s=1}^{n}a_{si}b_{js}=\sum_{s=1}^{n}(A^T)_{is}(B^T)_{sj}$$

$$\frac{\partial\varphi}{\partial X}=\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=A^T B^T$$

注意，定理4.4.7是这里的特殊情况。

定理4.4.10 若矩阵变量$X\in\mathbf{R}^{n\times n}$的行列式 $\det(X)\neq 0$，标量函数为

$$\varphi(X)=\det(X)$$

则有

$$\frac{\partial\varphi}{\partial X}\triangleq\left[\frac{\partial\varphi}{\partial x_{ij}}\right]=(\det(X))(X^{-1})^T$$

证明 设 x_{ij} 的代数余子式为 X_{ij}，则据行列式的展开定理有

$$\varphi(X)=\det(X)=\sum_{i=1}^{n}x_{ij}X_{ij}$$

故有

$$\frac{\partial\varphi}{\partial x_{ij}}=X_{ij},\quad i, j=1, 2, \cdots, n$$

$$\frac{\partial\varphi}{\partial X}=[X_{ij}]$$

而$[X_{ij}]\in\mathbf{R}^{n\times n}$是矩阵$X^T$的伴随矩阵，从而有

$$\frac{\partial\varphi}{\partial X}=[X_{ij}]=(\det(X))(X^{-1})^T$$

下面举几个应用的例子。

例 1（欧氏空间的线性拟合） 在m维欧氏空间 $\mathbf{R}^m$中，给定N个点

$$P_i=(x_1^{(i)}, x_2^{(i)}, \cdots, x_m^{(i)}),\quad i=1, 2, \cdots, N$$

欲求一超平面

$$H: K^Tx=k_0$$

其中 $K=(k_1, k_2, \cdots, k_m)^T$, $x=(x_1, x_2, \cdots, x_m)^T$ 使得各点P_i 到H的距离平方之和为最小。这样求得的H为点组 $\{P_1, P_2, \cdots, P_N\}$ 的最小二乘线性拟合。特别当$m=2$ 时，便是最常用的平面上点组的线性拟合。

解 很显然，P_i到H的距离的平方为

$$d_i^2=\frac{(K^TP_i-k_0)^2}{K^TK}\quad i=1, 2, \cdots, N$$

（它是二维情况下公式 $d=|ax+by+c|/\sqrt{a^2+b^2}$ 的推广）从而诸点 P_i 到H的距离平方之和

$$s=\sum_{i=1}^{N}d_i^2=\frac{1}{K^TK}\sum_{i=1}^{N}(K^TP_i-k_0)^2$$

很显然，超平面H:

$$K^T x = k_1 x_1 + k_2 x_2 + \cdots + k_m x_m = k_0$$

随着方程的系数k_1，k_2，…，k_m （它们构成的向量K为H的法向量）的改变而改变，所以s是K、k_0的函数，故令 $s = s(k_1, k_2, \cdots, k_m, k_0)$。我们的目的就是要确定一组$k_0$，$k_1$，…，$k_m$，使$s$达到极小。欲使$s$达到极小，其中的一个必要条件是

$$\frac{\partial s}{\partial k_0} = -\frac{2}{K^T K}\sum_{i=1}^{N}(K^T P_i - k_0) = 0$$

从而得到

$$k_0 = \frac{1}{N}\sum_{i=1}^{N} K^T P_i = K^T\left(\frac{1}{N}\sum_{i=1}^{N} P_i\right)$$

现在令

$$\bar{P} = \frac{1}{N}\sum_{i=1}^{N} P_i$$

它显然为N个给定点的几何中心。从而有

$$k_0 = K^T \bar{P}$$

$$H: \qquad K^T(x - \bar{P}) = 0$$

这表明所要求的最佳拟合超平面一定过中心$\bar{P}$。这时函数

$$s = \frac{1}{K^T K}\sum_{i=1}^{N}(K^T P_i - k_0)^2$$

$$= \frac{1}{K^T K}\sum_{i=1}^{N}[K^T(P_i - \bar{P})]^2$$

若令 $Y_i = P_i - \bar{P} \in \mathbf{R}^m$，$\sum_{i=1}^{N} Y_i Y_i^T = A \in \mathbf{R}^{m \times m}$，则据 $K^T Y_i = Y_i^T K$可得

$$s = \frac{1}{K^T K}\sum_{i=1}^{N}[K^T Y_i]^2$$

$$= \frac{1}{K^T K}\sum_{i=1}^{N} K^T Y_i K^T Y_i$$

$$= \frac{1}{K^T K} \sum_{i=1}^{N} K^T Y_i Y_i^T K$$

$$= \frac{1}{K^T K} K^T \sum_{i=1}^{N} Y_i Y_i^T K$$

$$= \frac{1}{K^T K} K^T A K$$

若再令 $Z = K / \| K \|$，$\| K \| = \sqrt{K^T K}$，则 $\| Z \| = 1$。这时，原来求多元函数 $s = K^T A K / K^T K$ 的极小点K的问题，就变成下述条件极值问题

$$\min_{\| Z \| = 1} Z^T A Z$$

引入Lagrange乘子 λ，化成无条件极值问题，即求

$$\phi(z,\ \lambda) \triangleq Z^T A Z + \lambda (Z^T Z - 1)$$

的极小点。为此，设（Z_0，λ_0）为该函数的极小点，据微分学的知识，并用矩阵标量函数的求导公式（定理4.4.8）便得到

$$\left.\frac{\partial \phi}{\partial Z}\right|_{(z_0,\ \lambda_0)} = 2 A Z_0 + 2 \lambda_0 Z_0 = 0$$

$$\left.\frac{\partial \phi}{\partial \lambda}\right|_{(z_0,\ \lambda_0)} = Z_0^T Z_0 - 1 = 0$$

由这二个方程可得

$$A Z_0 = -\lambda_0 Z_0, \qquad \| Z_0 \| = 1$$

且

$$\min S(K) = \lambda_{\min}$$

其中$\lambda_{\min}$ 为矩阵A（至少是半正定阵）的最小特征值。Z_0 为 A 的对应$\lambda_{\min}$ ($= -\lambda_0$)的单位特征向量。从而可以断言：

1. 要求的超平面H过$\bar{P}$；

2. 若$\lambda_{\min}$ 为A 的单特征值，则对应 $\lambda_{\min}$ 有唯一的独立的特征向量 Z_0，从而过$\bar{P}$ 且以Z_0为法方向的H是唯一的要求的超平面。

3. 若对应 λ_{min} 共有 r 个线性独立的特征向量 $\{Z_i\}$，则过 P 以 Z_i 或 $\{Z_i\}$ 的线性组合为法方向的 H 都是最佳拟合超平面，有无穷多。特别是当 $r=m$ 时，过 P 的任何 H 都是最佳拟合超平面。

例2 给定 $A_i \in \mathbf{R}^{m\times n}$，$b_i \in \mathbf{R}^m$，$i=1, 2, \cdots, k$。则 $x_0 \in \mathbf{R}^n$ 是函数

$$f(x)=\sum_{i=1}^{k} \| A_i x - b_i \|^2$$

的极小点的充分必要条件为：x_0 是方程

$$\left(\sum_{i=1}^{k} A_i^T A_i\right)x = \sum_{i=1}^{k} A_i^T b_i$$

的解，或方程组

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

的最小二乘解。

证明 必要性。由于

$$f(x)=\sum_{i=1}^{k} \| A_i x - b_i \|^2$$

$$=\sum_{i=1}^{k} (A_i x - b_i)^T (A_i x - b_i)$$

$$=\sum_{i=1}^{k} (x^T A_i^T A_i x - 2x^T A_i^T b_i + b_i^T b_i)$$

若 x_0 为 $f(x)$ 的极小点，则应有

$$\left.\frac{\partial f}{\partial x}\right|_{x_0} \triangleq \left[\frac{\partial f}{\partial x_i}\right]_{x_0} = \sum_{i=1}^{k} [2A_i^T A_i x_0 - 2A_i^T b_i] = 0$$

即

$$\left(\sum_{i=1}^{k} A_i^T A_i\right) x_0 = \sum_{i=1}^{k} A_i^T b_i$$

就是说x_0是方程

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_k \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

的法方程$\left(\sum_{i=1}^{k} A_i^T A_i\right)x = \sum_{i=1}^{k} A_i b_i$的解，当然是方程的最小二乘解。

充分性。x_0为方程(*)的最小二乘解，据定义它应该是函数$f(x) = \sum_{i=1}^{k} \| A_i x - b_i \|^2$的极小点。

例3 （约束最小二乘问题）给定 $A \in \mathbf{R}^{m \times n}$ 且是列满秩阵，$b \in \mathbf{R}^n$，$B \in \mathbf{R}^{k \times n}$，$d \in R(B)$。试求约束极小化问题

$$\min_{Bx=d} \| Ax - b \|^2$$

的解。也就是求函数$f(x) = \| Ax - b \|^2$在约束$Bx = d$下的极小点和极小值。

解 引入 Lagrange 乘子$2\lambda \in \mathbf{R}^k$，化成等价的无约束极值问题：令

$$\varphi(x, \lambda) = \| Ax - b \|^2 + 2\lambda^T(Bx - d)$$

若(x_0, λ_0)为$\varphi(x, \lambda)$的极点，则应有

$$\left.\frac{\partial \varphi}{\partial x}\right|_{(x_0, \lambda_0)} = 2A^T A x_0 - 2A^T b + 2B^T \lambda_0 = 0$$

$$\left.\frac{\partial \varphi}{\partial \lambda}\right|_{(x_0, \lambda_0)} = 2(Bx_0 - d) = 0$$

说明极点(x_0, λ_0)应满足方程

$$\begin{bmatrix} A^T A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ \lambda_0 \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix} \qquad (**)$$

注意到

$$\frac{\partial^2 f}{\partial x^2} \triangleq \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] = 2A^T A$$

是正定阵，故x_0必为$f(x)$的极小点。在方程(＊＊)两端左乘上满秩阵（即进行初等变换）

$$\begin{bmatrix} I_n & 0 \\ -B(A^TA)^{-1} & I_k \end{bmatrix}$$

便得到

$$\begin{bmatrix} A^TA & B^T \\ 0 & -B(A^TA)^{-1}B^T \end{bmatrix}\begin{bmatrix} x_0 \\ \lambda_0 \end{bmatrix}=\begin{bmatrix} A^Tb \\ d-B(A^TA)^{-1}A^Tb \end{bmatrix}$$

解该方程组便得

$$\lambda_0=[B(A^TA)^{-1}B^T]^{(1)}[B(A^TA)^{-1}A^Tb-d]$$

$$x_0=(A^TA)^{-1}(A^Tb-B^T\lambda_0)$$

$$f(x_0)=f((A^TA)^{-1}(A^Tb-B^T\lambda_0))$$

例4 在信息论中Shannon引入了熵的概念：随机向量$X\in\mathbf{R}^n$具有概率密度$p(X)$，其熵定义为

$$H=-E_x[\log p(X)]$$

其中$E_x[\varphi(X)]$表示对随机变量X的函数$\varphi(X)$取数学期望。由此建立了统计理论中极小化极大熵的估计方法。还证明了当X为高斯分布时，其熵$H=\frac{1}{2}\log(\det\sum)+\frac{1}{2}n(\log 2\pi+1)$达到极大值，其中$\sum=E_x[(X-\mu)(X-\mu)^T]$为$X$的协方差阵。在最佳线性无偏估计中，Cramer-Rao不等式的推导也要涉及到对矩阵的标量函数

$$\varphi(M)=\log\det(M)$$

的求导。为此，给出一个结论：

在矩阵集合

$$D=\{X\in\mathbf{R}^{n\times n}\mid X\text{是正定对称的}\}$$

上定义函数

$$f(X)=\log\det(X)$$

则有

$$\frac{\partial f}{\partial X}\triangleq\left[\frac{\partial f}{\partial x_{ij}}\right]=X^{-1}$$

证明 设 $X=(x_{ij})$ x_{ij}的代数余子式为X_{ij}，则对X的伴随矩阵$\mathrm{adj}X$来讲，有

$$(\mathrm{adj}X)_{ij}=X_{ji}, \qquad X(\mathrm{adj}X)=(\det X)I$$

$$(X\mathrm{adj}X)_{ii}=\sum_{k=1}^{n}x_{ik}X_{ik}=\det X, \quad i=1, 2, \cdots, n$$

上面最后一式两端同时取对数，便有

$$f(X)=\log(\det X)=\log\left(\sum_{k=1}^{n}x_{ik}X_{ik}\right)$$

再关于x_{ij}求导，便有

$$\frac{\partial f}{\partial x_{ij}}=\frac{X_{ij}}{\sum_{k=1}^{n}x_{ik}X_{ik}}=\frac{(\mathrm{adj}X)_{ji}}{\det X}, i=1, 2, \cdots, n$$

从而得到

$$\frac{\partial f}{\partial X}=\left[\frac{\partial f}{\partial x_{ij}}\right]=\frac{\mathrm{adj}X^T}{\det(X^T)}=(X^T)^{-1}=X^{-1}。$$

第五节 矩阵分析在微分方程中的应用

在数学或工程技术中，经常要研究一阶微分方程组：

$$\frac{dx_i(t)}{dt}=\sum_{j=1}^{n}a_{ij}(t)x_j(t)+g_i(t)$$

$$x_i(0)=c_i \qquad i=1, 2, \cdots, n$$

如果记

$$x(t)=(x_1(t), x_2(t), \cdots, x_n(t))^T$$

$$x(0)=(c_1, c_2, \cdots, c_n)^T$$

$$g(t)=(g_1(t), g_2(t), \cdots, g_n(t))^T$$

$$A(t)=[a_{ij}(t)]_{n\times n}$$

则上述方程组可写为

$$\frac{dx(t)}{dt}=A(t)x(t)+g(t)$$

$$x(0)=c$$

大家都很熟悉，如果$n=1$，$A(t)=a$，上面方程便呈

$$\dot{x}(t)=ax(t)+g(t)$$
$$x(0)=c$$

对方程移项，乘上积分因子e^{-at}，便有

$$\frac{d}{dt}(e^{-at}x(t))=e^{-at}\dot{x}(t)-ae^{-at}x(t)=e^{-at}g(t)$$

两边关于s在$[0,t]$上取积分，得到方程的解

$$x(t)=e^{at}c+\int_0^t e^{a(t-s)}g(s)ds$$

这是很简单的一个问题。现在我们面临的是一个用矩阵形式给出的方程组，能否作类似的处理呢？下面进行分析。如果$A(t)$是常数矩阵A，则方程呈

$$\frac{dx}{dt}=Ax(t)+g(t)$$
$$x(0)=c$$

能否象前面样，用一个积分因子去乘方程

$$\dot{x}(t)-Ax(t)=g(t)$$

的两端化成全微分的形式，然后取积分得到解呢？如果可以的话，这个积分因子是什么？稍作分析便知：前边之所以用积分因子e^{-at}能获得成功，是因为

1． e^{-at}使得$d[e^{-at}x(t)]=e^{-at}\dot{x}(t)-ae^{-at}x(t)$，实质上是因为$\frac{d}{dt}e^{-at}=-ae^{-at}$；

2． 对任意的$t\in\mathbf{R}$，$e^{-at}\neq0$故积分后除以e^{-at}得到$x(t)=e^{at}c+\int_0^t e^{a(t-s)}g(s)ds$。

这样分析之后，便很自然地想到取积分因子e^{-At}，因为我们能够看出下列事实：

第一，级数$\sum_{n=0}^{\infty}\frac{(-tA)^n}{n!}$对任意的$A\in\mathbf{R}^{n\times n}$及$t\in\mathbf{R}$都是收敛的，并且可记为$e^{-At}=\sum_{n=0}^{\infty}\frac{(-At)^n}{n!}$；

第二，e^{-tA}与A是可交换的，即$A\cdot e^{-tA}=e^{-tA}\cdot A$；

第三，$\frac{d}{dt}e^{-tA}=\frac{d}{dt}\left(\sum_{n=0}^{\infty}\frac{(-tA)^n}{n!}\right)=-Ae^{-tA}$；

第四，对任意的$t\in\mathbf{R}$及$A\in\mathbf{R}^{n\times n}$，$\det(e^{-tA})\neq 0$，即$e^{-tA}$可逆，且$e^{-tA}=(e^{tA})^{-1}$。

前三条是显然的，而第四条包含在下面的一个定理之中。

定理4.5.1 设A，$B\in\mathbf{C}^{n\times n}$，$t\in\mathbf{R}$，则有

1. $\det(e^{At})=e^{[\mathrm{tr}(A)]t}$
2. $(e^{tA})^{-1}=e^{-tA}$
3. 若$AB=BA$，便有

$$e^{(A+B)t}=e^{At}\cdot e^{Bt}=e^{Bt}\cdot e^{At}$$

证明 若A的特征值为λ_1，λ_2，…，λ_n，则tA的特征值为$\lambda_1 t$，$\lambda_2 t$，…，$\lambda_n t$，而e^{tA}的特征值（据推论4.3.10）为$e^{\lambda_1 t}$，e^{λ_2}，…，$e^{\lambda_n t}$。从而据矩阵的行列式与特征值之间的关系可得

$$\det(e^{tA})=\prod_{i=1}^{n}e^{\lambda_i t}=e^{\sum_{i=1}^{n}\lambda_i t}=e^{t(\mathrm{tr}A)}$$

由于$e^{t(\mathrm{tr}A)}\neq 0$，故$e^{At}$总是可逆的。这便证明了1。根据定义及$AB=BA$，可得到

$$e^{(A+B)t}=\sum_{n=0}^{\infty}\frac{t^n(A+B)^n}{n!},\quad (A+B)^nA=A(A+B)^n$$

$$B(A+B)^n=(A+B)^nB,\qquad n=1,2,\cdots,$$

现构造一个函数矩阵

$$C(t)=e^{(A+B)t}\cdot e^{-At}\cdot e^{-Bt}$$

两端关于t求导便有

$$C'(t)=(A+B)e^{(A+B)t}\cdot e^{-At}\cdot e^{-Bt}$$
$$-e^{(A+B)t}\cdot Ae^{-At}\cdot e^{-Bt}-e^{(A+B)t}\cdot e^{-At}Be^{-Bt}$$
$$=(A+B-A-B)e^{(A+B)t}\cdot e^{-At}\cdot e^{-Bt}=0$$

这表明$C(t)$是常数矩阵，故有

$$C(1)=C(0)=e^0\cdot e^0\cdot e^0=I$$

亦即

$$e^{A+B}\cdot e^{-A}\cdot e^{-B}=I$$

同理可证

$$e^{A+B}\cdot e^{-B}\cdot e^{-A}=I$$

特别取 $B=-A$，从上面两式可得

$$e^A\cdot e^{-A}=I, \quad e^{-B}\cdot e^B=I$$

这便证明了$(e^A)^{-1}=e^{-A}$，$(e^B)^{-1}=e^{-B}$，从而有

$$e^{A+B}=e^A\cdot e^B=e^B\cdot e^A$$

顺便也证明了，对任意的$t\in\mathbf{R}$，及A，$B\in\mathbf{C}^{n\times n}$有

$$e^{-At}=(e^{tA})^{-1}$$
$$e^{(A+B)t}=e^{At}\cdot e^{Bt}=e^{Bt}\cdot e^{At}, \qquad (AB=BA)$$

值得注意的是，当$AB\neq BA$时，$e^{A+B}=e^A\cdot e^B$或$e^{A+B}=e^B\cdot e^A$可以是不成立的。例如我们取

$$A=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}, \qquad B=\begin{pmatrix}1 & -1\\ 0 & 0\end{pmatrix}$$

则有 $e^A=I+(e-1)A$，$e^B=I+(e-1)B$

$$e^A\cdot e^B=\begin{pmatrix}e^2 & -(e-1)^2\\ 0 & 1\end{pmatrix}, \qquad e^{A+B}=\begin{pmatrix}e^2 & 0\\ 0 & 1\end{pmatrix}$$

有了上述的准备，常系数微分方程组

$$\dot{x}(t)=Ax(t)+g(t)$$
$$x(0)=c$$

就可以象标量方程一样求解了。因为有

$$\frac{d}{dt}(e^{-At}x(t))=e^{-At}(\dot{x}(t)-Ax(t))=e^{-At}g(t)$$

故两端关于 s 在[0，t]上取积分可得

$$e^{-At}x(t)-x(0)=\int_0^t e^{-As}g(s)ds$$

两端同时乘矩阵e^{At}，便得到方程的解

$$x(t)=e^{At}x(0)+\int_0^t e^{A(t-s)}g(s)ds$$

很容易验证$x(t)$是方程的解。至于给定 A，e^{At}如何计算，后面再作介绍。

对于一般的 $A(t)$，能否有一个类似积分因子的函数矩阵 $F(t)$，具有如下性质：

$$\dot{F}(t)=-F(t)A(t)，\det(F(t))\neq0，F(0)=I$$

如果存在，则在方程

$$\dot{x}(t)-A(t)x(t)=g(t)$$

两端左乘$F(t)$，便有

$$\frac{d}{dt}[F(t)x(t)]=F(t)\dot{x}(t)-F(t)A(t)x(t)$$

$$=F(t)g(t)$$

再关于 t 在[0，t]上取积分，便得

$$F(t)x(t)-F(0)x(0)=\int_0^t F(s)g(s)ds$$

从而求得方程的解

$$x(t)=F^{-1}(t)x(0)+\int_0^t F^{-1}(t)F(s)g(s)ds \qquad (*)$$

现在的问题是：给定了$A(t)$，如何构造$F(t)$？$F(t)$还有哪些性质？这些问题都比较复杂。我们打算按下述步骤考虑方程的求解问题。首先考虑齐次方程 $\dot{x}=A(t)x(t)$ 的基础解系，或基础解阵的构造；然后证明齐次方程满足初始条件的解是唯一的；最后求非齐次方程 $\dot{x}=A(t)x(t)+g(t)$ 满足初始条件的解。

对于线性齐次方程组的基础解系，一般是这样定义的：第

一，基础解系是一组特解；第二，这组特解是线性无关的；第三，方程组的每个解皆可用之线性表出。在前面推导的非齐次方程的解（＊）中，令$g(s)=0$，则得齐次方程的解

$$x(t)=F^{-1}(t)x(0)$$

若把$F^{-1}(t)$进行列分块：

$$F^{-1}(t)=[Y_1(t), Y_2(t), \cdots, Y_n(t)]$$

并注意到$x(0)=(c_1, c_2, \cdots, c_n)^T$，则有

$$X(t)=c_1Y_1(t)+c_2Y_2(t)+\cdots+c_nY_n(t)$$

从而说明齐次方程 $\dot{x}=A(t)x(t)$ 的满足任意初始条件的特解都可以用$\{Y_1(t), Y_2(t), \cdots, Y_n(t)\}$线性表出；由于$F^{-1}(t)$是可逆的，所以$\{Y_1(t), Y_2(t), \cdots, Y_n(t)\}$是线性无关的；据假设$\dot{F}(t)=-F(t)A(t)$，从而对式$F^{-1}(t)F(t)=I$两端求导便得

$$\frac{d}{dt}(F^{-1}(t))=-F^{-1}(t)\dot{F}(t)F^{-1}(t)$$

$$=A(t)F^{-1}(t)$$

亦即

$$\frac{dY_i}{dt}=A(t)Y_i(t) \qquad , i=1, 2, \cdots, n$$

这表明$F^{-1}(t)$就是由齐次方程的一个基础解系构成的。前面所说的寻找积分因子就等价于寻找基础解阵。现在给出定义：

定义4.5.2 对于给定的$A(t)$，如果存在n阶函数矩阵$X(t)$满足

$$\frac{dX(t)}{dt}=A(t)X(t)$$

$$X(0)=I$$

则称$X(t)$为微分方程$\dot{x}=A(t)x(t)$的基础解阵，或简称关于$A(t)$的基础解阵。

据定义可以看出，$X(t)$的每个列向量都是方程$\dot{x}=A(t)x(t)$的解。而方程的满足初始条件$x(0)=c$的特解也都可以用$X(t)$

的列向量线性表示为$x(t)=X(t)c$。关于$X(t)$是否可逆，有下面定理保证。

定理4.5.3（Jacobi恒等式） 若$A(t)$在$[0, b]$上连续且它的基础解阵$X(t)$存在，则$X(t)$可逆且

$$\det(X(t))=e^{\int_0^t tr(A(s))ds}$$

证明 设$A(t)=(a_{ij}(t))$并对$X(t)$进行行分块：

$$X(t)=[x_{ij}(t)]=\begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_n(t) \end{bmatrix}$$

其中$X_i(t)=(x_{i1}(t), x_{i2}(t), \cdots, x_{in}(t))$。则据

$$\frac{dX(t)}{dt}=A(t)X(t)$$

可得

$$\frac{dX_j(t)}{dt}=[A(t)X(t)]_j$$

$$=\sum_{k=1}^{n}a_{jk}(t)X_k(t), \quad j=1, 2, \cdots, n$$

据行列式的求导法则有

$$\frac{d}{dt}[\det X(t)]=\sum_{i=1}^{n}\det\left(\begin{bmatrix} X_1(t) \\ \vdots \\ X_{i-1}(t) \\ \frac{d}{dt}X_i(t) \\ \vdots \\ X_n(t) \end{bmatrix}\right)$$

$$
= \sum_{i=1}^{n} \det\left(\begin{pmatrix} X_1(t) \\ \vdots \\ X_{i-1}(t) \\ \sum_k a_{ik}(t) X_k(t) \\ \vdots \\ X_n(t) \end{pmatrix} \right)
$$

$$
= \sum_{i=1}^{n} a_{ii}(t)\det(X(t))
$$

$$
= \mathrm{tr}(A(t))\det(X(t))
$$

令 $\varphi(t)=\det(X(t))$，并注意到 $\varphi(0)=\det(X(0))=1$，说明 $\varphi(t)$在 $t=0$ 的充分小邻域内不等于0，故有

$$
\int_0^t \frac{d\varphi}{\varphi} = \int_0^t \mathrm{tr}(A(s))ds
$$

从而得到

$$
\det(X(t)) = \varphi(t) = e^{\int_0^t \mathrm{tr}(A(s))ds}
$$

通过上面的讨论，可以看出方程组

$$
\dot{x} = A(t)x(t)
$$

$$
x(0) = c
$$

的解的存在唯一性问题等价于基础解阵 $X(t)$ 的存在唯一性问题。下面首先证明$X(t)$的存在性。

定理4.5.4 如果 $A(t)=(a_{ij}(t))$对每个$t\in\Omega$都连续，则对每个 $t\in\Omega$ 都存在它的基础解阵 $X(t)$，也就是存在n阶函数矩阵$X(t)$满足

$$
\frac{dX}{dt} = A(t)X(t)
$$

$$
X(0) = I
$$

证明 对每个给定 $\tau\in\Omega$，$A(\tau)$是一个常数矩阵，我们规定它的范数为

$$\| A(\tau) \| = \sum_i \sum_j |a_{ij}(\tau)|$$

并对任意的$t \in \Omega$，定义函数

$$\mu(t) = \sup\{ \| A(\tau) \| \mid |\tau| \leqslant |t| \}$$
$$= \sup\{ \sum_i \sum_j |a_{ij}(\tau)| \mid |\tau| \leqslant |t| \}$$

显然，当s，$t \in \Omega$且满足$|s| \leqslant |t|$时有$\mu(s) \leqslant \mu(t)$。现在通过构造矩阵序列来获得$X(t)$，为此令

$$X_0(t) = I$$

$$X_1(t) = I + \int_0^t A(s) X_0(s) ds$$

$$X_2(t) = I + \int_0^t A(s) X_1(s) ds$$

$$\vdots$$

$$X_{n+1}(t) = I + \int_0^t A(s) X_n(s) ds$$

然后，证明序列$\{X_n(t)\}$是收敛的，并证明它的极限$X(t)$就是我们要得到的$X(t)$。我们首先建立一个不等式，通过它证明$\{X_n(t)\}$是一个Cauchy列。下面先就$t \geqslant 0$且$t \in \Omega$来证，这时便有

$$\| X_1(t) - X_0(t) \| = \left\| \int_0^t A(s) ds \right\|$$

$$= \sum_i \sum_j \left| \left(\int_0^t A(s) ds \right)_{ij} \right|$$

$$= \sum_i \sum_j \left| \int_0^t a_{ij}(s) ds \right|$$

$$\leqslant \sum_i \sum_j \int_0^t |a_{ij}(s)| ds$$

$$= \int_0^t \left(\sum_i \sum_j |a_{ij}(s)| \right) ds$$

$$=\int_0^t \|A(s)\| ds$$

$$\leq \mu(t)\int_0^t ds = t\mu(t)$$

$$\|X_2(t)-X_1(t)\| = \left\|\int_0^t A(s)[X_1(s)-X_0(s)]ds\right\|$$

$$\leq \int_0^t \|A(s)\| \, \|X_1(s)-X_0(s)\| ds$$

$$\leq \int_0^t \mu(s)\cdot s\mu(s) ds$$

$$\leq \mu^2(t)\int_0^t s ds = \frac{1}{2!}t^2\mu^2(t)$$

$$\|X_3(t)-X_2(t)\|$$

$$= \left\|\int_0^t A(s)[X_2(s)-X_1(s)]ds\right\|$$

$$\leq \int_0^t \|A(s)\| \, \|X_2(s)-X_1(s)\| ds$$

$$\leq \int_0^t \mu(s)\frac{1}{2!}s^2\mu^2(s) ds$$

$$\leq \mu^3(t)\int_0^t \frac{1}{2!}s^2 ds = \frac{1}{3!}t^3\mu^3(t)$$

依此类推便可证明，对一切$0\leqslant t\in\Omega$及$n\geqslant 0$都有

$$\|X_{n+1}(t)-X_n(t)\| \leqslant \frac{t^{n+1}\mu^{n+1}(t)}{(n+1)!}$$

类似地对 $0>t\in\Omega$，也可证明

$$\|X_{n+1}(t)-X_n(t)\| \leqslant \frac{(-t)^{n+1}\mu^{n+1}(t)}{(n+1)!}$$

结合一起，对一切$t\in\Omega$及$n\geqslant 0$有

$$\| X_{n+1}(t) - X_n(t) \| \leqslant \frac{|t|^{n+1}(\mu(t))^{n+1}}{(n+1)!}$$

据此再利用范数公理中的三角不等式，对任意的正整数 l，都有

$$\begin{aligned} &\| X_{n+l}(t) - X_n(t) \| \\ &= \left\| \sum_{j=n}^{n+l-1} (X_{j+1}(t) - X_j(t)) \right\| \\ &\leqslant \sum_{j=n}^{n+l-1} \| X_{j+1}(t) - X_j(t) \| \\ &\leqslant \sum_{j=n}^{n+l-1} \frac{[|t|\mu(t)]^{j+1}}{(j+1)!} = \sum_{j=n+1}^{n+l} \frac{[|t|\mu(t)]^j}{j!} \end{aligned}$$

由于右端和式中的每一项皆为正函数，故对一切$m>n$，都有

$$\| X_m(t) - X_n(t) \| \leqslant \sum_{j=n+1}^{\infty} \frac{[|t|\mu(t)]^j}{j!}$$

根据e^x的 Taylor 级数展式，可知

$$e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!} = \sum_{j=0}^{n} \frac{x^j}{j!} + e^{\theta} \frac{x^{n+1}}{(n+1)!}, \quad 0<\theta<x$$

并且当$x>0$时，e^x是增函数，故又有$e^\theta<e^x$。从而前面的推导结果可以写为

$$\begin{aligned} \| X_m(t) - X_n(t) \| &\leqslant \sum_{j=n+1}^{\infty} \frac{[|t|\mu(t)]^j}{j!} \\ &\leqslant e^{|t|\mu(t)} \frac{[|t|\mu(t)]^{n+1}}{(n+1)!} \qquad (*) \end{aligned}$$

由于$A(t)$在Ω上连续，故对任意的$t\in\Omega$，都有$\mu(t)<\infty$，从而$e^{|t|\mu(t)}<\infty$。因此根据

$$\lim_{n\to\infty} \frac{[|t|\mu(t)]^{n+1}}{(n+1)!} = 0$$

便可推出，当m，$n\to\infty$时，$\| X_m(t) - X_n(t) \| \to 0$。这说明序列$\{X_n(t)\}$是一个 Cauchy 列，并且是一致收敛的。据定义可知，

每个$X_n(t)$都在Ω上连续，故其极限

$$X(t)=\lim_{n\to\infty}X_n(t)$$

也是连续的。再根据$X_{n+1}(t)$的定义：

$$X_{n+1}(t)=I+\int_0^t A(s)X_n(s)ds \qquad (**)$$

上式两端取极限，并利用一致收敛的性质，便有

$$X(t)=I+\int_0^t A(s)X(s)ds$$

这表明我们通过迭代格式(**)所得到的 $X(t)$ 满足上面这个积分方程。如果该方程两端令 $t=0$ 代入，便有$X(0)=I$；如果该方程两端关于 t 求导，便又有

$$\frac{dX}{dt}=A(t)X(t)$$

从而证明了 $X(t)$ 便是 $A(t)$ 的一个基础解阵，当然也证明了$X(t)$的存在性。

值得指出的是：定理 4.5.4 是一个构造性的证明，它给出了一个逼近$X(t)$的迭代程序。由于对一般的 $A(t)$ 是求不出$X(t)$的精确表达，所以求其近似的方法是很重要的。当我们关于不等式(*)求极限（$m\to\infty$，n 固定）时，便有

$$\|X(t)-X_n(t)\|\leqslant e^{|t|\mu(t)}\frac{[|t|\mu(t)]^{n+1}}{(n+1)!}$$

该式给出了第 n 次迭代的误差估计，并且可以根据预先要求的精度，估计需要迭代的次数 n。

矩阵$A(t)$的基础解阵$X(t)$的存在，意味着方程 $\dot{x}=A(t)\cdot x(t)$ 解的存在，而且满足初始条件 $x(0)=c$ 的解就是 $x(t)=X(t)c$。如果我们还能够证明 $A(t)$ 的基础解阵是唯一的话，则很容易推出方程

$$\dot{x}=A(t)x(t), \qquad x(0)=c$$

的解也是唯一的。因为若有两个解 $x(t)$，$y(t)$，则据$X(t)$的唯

一性及 $x(t)$，$y(t)$ 皆可用 $X(t)$ 表出，即 $x(t)=X(t)c_1$，$y(t)=X(t)c_2$，且满足初始条件$x(0)=y(0)=c$，则便有

$$0=x(0)-y(0)=X(0)c_1-X(0)c_2=c_1-c_2$$

这说明$c_1=c_2$，即 $x(t)=y(t)$。现在转而讨论$A(t)$的基础解阵$X(t)$的唯一性。如果 $X(t)$，$Y(t)$ 皆为 $A(t)$ 的基础解阵，令$Z(t)=X(t)-Y(t)$，便有

$$\begin{aligned}\frac{dZ}{dt}&=\frac{dX}{dt}-\frac{dY}{dt}\\&=A(t)X(t)-A(t)Y(t)\\&=A(t)Z(t)\end{aligned}$$

$$Z(0)=X(0)-Y(0)=I-I=0$$

这表明$Z(t)$满足

$$\frac{dZ}{dt}=A(t)Z(t)$$

$$Z(0)=0$$

因此，如果我们能证明上述方程的解$Z(t)$对一切$t\in\Omega$都有 $Z(t)=0$，即 $Z(t)=X(t)-Y(t)=0$，也就证明了基础解阵的唯一性。

定理4.5.5 如果$A(t)$在Ω上连续，且$Z(t)$满足

$$\frac{dZ}{dt}=A(t)Z(t) \qquad t\in\Omega$$

$$Z(0)=0$$

则有 $Z(t)=0$。

证明 上面的微分方程等价于积分方程

$$Z(t)=\int_0^t A(s)Z(s)ds, \qquad t\in\Omega$$

仿照定理 4.5.4 的证明，令

$$\xi(t)=\sup\{\|Z(\tau)\| \mid |\tau|\leqslant|t|\}$$

则当$|s|\leqslant|t|$时有$\xi(s)\leqslant\xi(t)$。仍设$0\leqslant t\in\Omega$，便有

$$\|Z(t)\|=\left\|\int_0^t A(s)Z(s)ds\right\|$$

$$\leqslant\int_0^t \|A(s)\|\ \|Z(s)\|ds$$

$$\leqslant\int_0^t \mu(s)\zeta(s)ds\leqslant t\mu(t)\zeta(t)$$

$$\|Z(t)\| = \left\|\int_0^t A(s)Z(s)ds\right\|$$

$$\leqslant\int_0^t \|A(s)\|\ \|Z(s)\|ds$$

$$\leqslant\int_0^t \mu(s)s\mu(s)\zeta(s)ds$$

$$\leqslant\mu^2(t)\zeta(t)\int_0^t sds = \frac{\mu^2(t)\zeta(t)t^2}{2!}$$

依此类推，便推出对一切n及$0\leqslant t\in\Omega$有

$$\|Z(t)\|\leqslant\frac{\mu^n(t)\zeta(t)t^n}{n!}$$

同样的方法，可推出对一切n及$0>t\in\Omega$也有

$$\|Z(t)\|\leqslant\frac{\mu^n(t)\zeta(t)(-t)^n}{n!}$$

合在一起，对一切n及$t\in\Omega$有

$$\|Z(t)\|\leqslant\frac{\mu^n(t)\zeta(t)|t|^n}{n!}\longrightarrow 0\ (n\to\infty)$$

从而证明了$\|Z(t)\|=0$，即$Z(t)=0$，由于$t\in\Omega$的任意性，可知$Z(t)=0$，$t\in\Omega$。

迄今为止，我们完成了方程

$$\dot{x}=A(t)x(t),\qquad x(0)=c$$

的解的存在与唯一性证明，并且迭代序列

$$X_0(t)=I$$

$$X_{n+1}(t)=I+\int_0^t A(s)X_n(s)ds,\ n=0,1,2,\cdots$$

给出了逼近基础解阵 $X(t)$ 具体可行的方法。也很容易看出，当 $A(t)=A$时，上面的迭代序列为

$$X_0(t)=I,\qquad X_{n+1}(t)=\sum_{k=0}^{n+1}\frac{t^kA^k}{k!}\quad n=0,1,2,\cdots$$

说明常系数的一阶线性常微分方程组 $\dot{x}=Ax(t)$ 的基础解阵就是 e^{tA}。

例 假定在方程 $\dot{x}=A(t)x(t), x(0)=c$ 中取 $A(t)=\begin{bmatrix} t & 1\\ 2 & t\end{bmatrix}$，$x(0)=(1,-1)^T$，求方程的近似解，并就 $|t|\leqslant 0.1$ 的情况，估计近似解与精确解之间的误差。

解 先算出基础解阵 $X(t)$ 的逼近序列：

$$X_0(t)=I$$

$$X_1(t)=I+\int_0^t\begin{bmatrix} s & 1\\ 2 & s\end{bmatrix}ds=I+\begin{bmatrix} \dfrac{t^2}{2} & t\\ 2t & \dfrac{t^2}{2}\end{bmatrix}$$

$$X_2(t)=I+\int_0^t\begin{bmatrix} 3s+s^3/2 & 1+3s^2/2\\ 2+3s^2 & 3s+s^3/2\end{bmatrix}ds$$

$$=\begin{bmatrix} 1+3t^2/2+t^4/8 & t+t^3/2\\ 2t+t^3 & 1+3t^2/2+t^4/8\end{bmatrix}$$

我们估计第二次迭代得到的近似解 $X_2(t)c$ 与精确解 $X(t)c$ 之间的误差：

$$\|x(t)-X_2(t)c\|=\|(X(t)-X_2(t))c\|$$
$$\leqslant\|X(t)-X_2(t)\|\,\|c\|$$

因为 $\|c\|=2$（这里取向量的范数为 $\|x\|=\sum|\xi_i|$）而 $\|X(t)-X_2(t)\|\leqslant e^{|t|\mu(t)}\dfrac{|t|^3\mu^3(t)}{3!}$，故有估计式

$$\|x(t)-X_2(t)c\|\leqslant\frac{2}{3!}|t|^3\mu^3(t)e^{|t|\mu(t)}.$$

又因为当 $t=0.1$ 时，可算得

$$
\begin{aligned}
\mu(0.1)&=\sup\{\,\|A(t)\| \mid |t|\leqslant 0.1\,\}\\
&=\sup\{\,2|t|+3 \mid |t|\leqslant 0.1\,\}=3.2
\end{aligned}
$$

从而当 $|t|\leqslant 0.1$ 时有

$$
\begin{aligned}
\|x(t)-X_2(t)c\| &\leqslant \frac{1}{3}e^{0.32}(0.32)^3\\
&<\frac{1}{3}\times 2\times\left(\frac{1}{3}\right)^3\\
&=\frac{2}{81}<0.025
\end{aligned}
$$

这时的近似解

$$
\bar{x}(t)=X_2(t)c=\begin{bmatrix} 1-t+3t^2/2-t^3/2+t^4/8 \\ -1+2t-3t^2/2+t^3-t^4/8 \end{bmatrix}
$$

下面考虑方程 $\dot{x}=A(t)x(t)+g(t)$，$x(0)=c$ 的解。和标量的常微分方程一样，采用常数变易法求解。设该方程的解为

$$
x=X(t)y(t)
$$

其中 $X(t)$ 为对应齐次方程的基础解阵，$y(t)$ 是待定的函数。故据基础解阵满足方程

$$
\dot{X}(t)=A(t)X(t),X(0)=I
$$

可得

$$
\begin{aligned}
\dot{x}&=\dot{X}(t)y(t)+X(t)\dot{y}(t)\\
&=A(t)X(t)y(t)+X(t)\dot{y}(t)\\
&=A(t)x+X(t)\dot{y}(t)
\end{aligned}
$$

将上式代入方程 $\dot{x}=A(t)x(t)+g(t)$ 的左端，并与右端比较消去相同的项，便得

$$
X(t)\dot{y}(t)=g(t)
$$

$$
\frac{dy}{dt}=X^{-1}(t)g(t)
$$

$$\int_0^t dy=\int_0^t X^{-1}(s)g(s)ds$$

$$y(t)-y(0)=\int_0^t X^{-1}(s)g(s)ds$$

据所设$x(t)=X(t)y(t)$，故有

$$c=x(0)=X(0)y(0)=Iy(0)=y(0)$$

$$y(t)=x(0)+\int_0^t X^{-1}(s)g(s)ds$$

$$\begin{aligned}x(t)&=X(t)y(t)\\&=X(t)x(0)+X(t)\int_0^t X^{-1}(s)g(s)ds\end{aligned}$$

这样就完成了非齐次方程的解过程。

第六节　稳定性与Lyapunov定理

很多物理系统都可用一阶微分方程组

$$\frac{dx}{dt}=A(t)x(t)$$
$$x(0)=c \qquad (*)$$

来描述。上一节的讨论表明，若$X(t)$为$A(t)$的基础解阵，则方程的唯一解为$x(t)=X(t)c$，它既是t的函数，同时也依赖于初始状态$x(0)=c$。所以我们可以通过初始状态c的摄动来研究它对解$x(t)$的影响，从而确定系统的稳定性。为了强调初始状态c对解$x(t)$的影响作用，特记$x_c(t)=X(t)c$，而把$x_0(t)=X(t)0=0$叫做方程的零解。本节旨在用前边学过的理论与方法研究方程所描述的物理系统的稳定性问题。

定义4.6.1 方程（*）的零解在Lyapunov意义下是稳定的，当且仅当对任意的$\varepsilon>0$，都存在$\delta(\varepsilon)>0$，使得对一切$t\geqslant0$，$\|c\|\leqslant\delta(\varepsilon)$隐含着

$$\| X_c(t) \| \leqslant \varepsilon$$

定义4.6.2 若方程（*）所描述的系统在Lyapunov意义下是稳定的，并且对从任意初始状态 c 出发的运动 $x_c(t)$，在 $t \to +\infty$ 都收敛到0，即

$$\lim_{t \to +\infty} x_c(t) = 0, \qquad \forall c \in \mathbf{C}^n$$

则称方程（*）的零解是渐近稳定的。

例1 方程 $\frac{dx}{dt} = x$，$x(0) = c$的零解是不稳定的。

因为$x_c(t) = e^t c$，对给定的$\varepsilon > 0$找不到只依赖于ε的$\delta(\varepsilon)$，使得$\| c \| \leqslant \delta(\varepsilon)$隐含$\| x_c(t) \| \leqslant \varepsilon$。

例2 方程 $\frac{dx}{dt} = x\sin t$的零解是稳定的。

因为对一切 t 都有$\| x_c(t) \| = \| e^{-\cos t} c \| \leqslant e \| c \|$。但却不是渐近稳定的，因为极限$\lim\limits_{t \to \infty} e^{-\cos t} c$不存在。

例3 方程 $\frac{dx}{dt} = -x$的零解是渐近稳定的。

因为$\| x_c(t) \| = \| e^{-t} c \| = e^{-t} \| c \| \longrightarrow 0 (t \to +\infty)$。

定理4.6.3 方程（*）的零解在Lyapunov意义下是稳定的，当且仅当$A(t)$的基础解阵是一致有界的，即对一切 $t \geqslant 0$，存在常数$\gamma > 0$，使得

$$\| X(t) \| \leqslant \gamma$$

证明 充分性：因为$x_c(t) = X(t)c$，所以对一切 $t \geqslant 0$有

$$\| x_c(t) \| = \| X(t)c \| \leqslant \| X(t) \| \| c \| \leqslant \gamma \| c \|$$

故若取$\delta = \varepsilon / \gamma$，便有$\| c \| \leqslant \delta$隐含着

$$\| x_c(t) \| \leqslant \varepsilon$$

必要性：若$x_0(t)$是稳定的而$X(t)$不是一致有界，则$X(t)$至少有一个元素当 $t \to +\infty$时变得任意大，不妨设这个元素为$x_{ij}(t)$。于是取初始状态 $c = (0, 0, \cdots, \alpha, 0, 0, \cdots, 0)^T$，其第 j 个分量α不为零，其余分量皆为零。这时$x_c(t)$的第 j 个分

量便是 $\alpha x_{ij}(t)$，显然不管 α 取得如何小，当 $t \longrightarrow +\infty$ 时，$\alpha x_{ij}(t)$ 都会任意大，因而零解是不稳定的，这与假设矛盾，故若零解是稳定的，$X(t)$ 必是一致有界的。

定理4.6.4 方程（$*$）的零解是渐近稳定的，其充分必要条件是

$$\lim_{t\to+\infty} X(t)=0$$

证明 充分性是显然的。必要性仿照定理4.6.3用反证法，读者自己完成。

当方程($*$)中的 $A(t)$ 为常数矩阵时，其解有明显的表达式

$$x_c(t)=e^{tA}c$$

所以它的稳定性有著名的 Lyapunov 判据。

定理4.6.5（Lyapunov 判据） 若 A 是常数矩阵，则方程 $\frac{dx}{dt}=Ax$ 的零解是渐近稳定的，其充分必要条件是 A 的所有特征值都有负实部。

证明 当 $A=PJP^{-1}$，$J=\mathrm{diag}(J_{n_1}(\lambda_1), J_{n_2}(\lambda_2), \cdots, J_{n_k}(\lambda_k))$ 时，有 $t^mA^m=P\mathrm{diag}(t^mJ^m_{n_1}(\lambda_1), t^mJ^m_{n_2}(\lambda_2), \cdots, t^mJ^m_{n_k}(\lambda_k))P^{-1}$。从而有

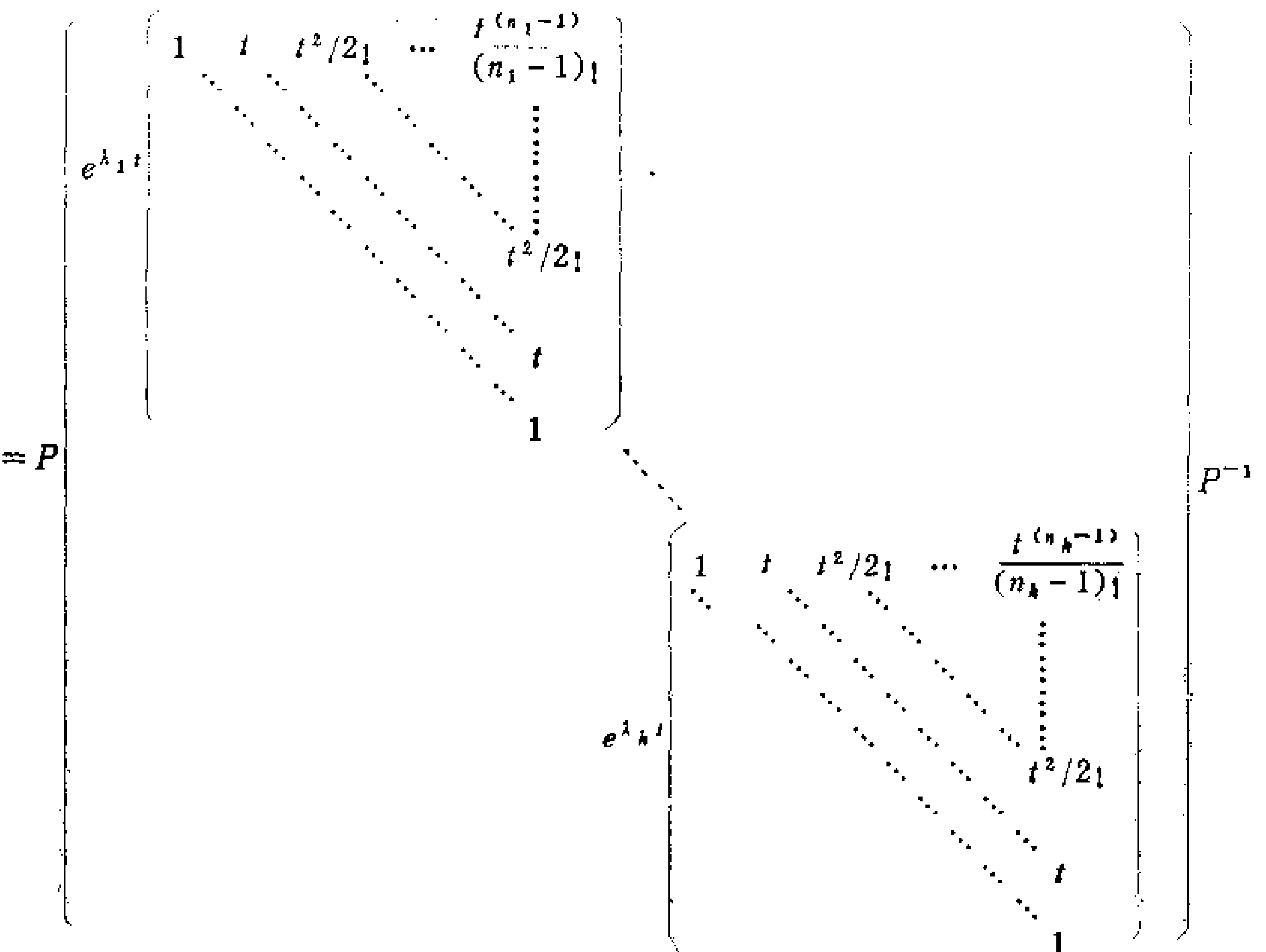
$$e^{tA}=\sum_{m=0}^{\infty}\frac{(tA)^m}{m!}=P\Big[\bigoplus_{i=1}^{k}\sum_{m=0}^{\infty}\frac{t^mJ^m_{n_i}(\lambda_i)}{m!}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{m=0}^{\infty}\frac{t^m}{m!}\sum_{j=0}^{n_i-1}c_m^j\lambda_i^{m-j}U^j_{n_i}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}U^j_{n_i}\sum_{m=j}^{\infty}\frac{t^m}{m!}\frac{m!}{(m-j)!j!}\lambda_i^{m-j}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}\sum_{j=0}^{n_i-1}\frac{t^jU^j_{n_i}}{j!}\sum_{m=j}^{\infty}\frac{(t\lambda_i)^{m-j}}{(m-j)!}\Big]P^{-1}$$

$$=P\Big[\bigoplus_{i=1}^{k}e^{\lambda_it}\sum_{j=0}^{n_i-1}\frac{(tU_{n_i})^j}{j!}\Big]P^{-1}$$



$$
= P \begin{pmatrix} e^{\lambda_1 t}\begin{pmatrix} 1 & t & t^2/2! & \cdots & \dfrac{t^{(n_1-1)}}{(n_1-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & t^2/2! \\ & & & \ddots & t \\ & & & & 1 \end{pmatrix} & & \\ & \ddots & \\ & & e^{\lambda_k t}\begin{pmatrix} 1 & t & t^2/2! & \cdots & \dfrac{t^{(n_k-1)}}{(n_k-1)!} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & t^2/2! \\ & & & \ddots & t \\ & & & & 1 \end{pmatrix} \end{pmatrix} P^{-1}
$$

设 $\lambda_l=\alpha_l+i\beta_l$，则 $e^{\lambda_l t}=e^{\alpha_l t}(\cos\beta_l t+i\sin\beta_l t)$，代入上式后可以看出 $P^{-1}e^{tA}P$ 的每个元素都具有如下的形式：$t^j e^{\alpha_l t}(\cos\beta_l t+i\sin\beta_l t)$。因此，当 $t\to+\infty$时，有

$$e^{tA}\longrightarrow 0 \Longleftrightarrow P^{-1}e^{tA}P\longrightarrow 0 \Longleftrightarrow \alpha_l<0$$

再据定理4.6.7知$\frac{dx}{dt}=Ax$ 的零解是渐近稳定的，当且仅当 $\alpha_l=\mathrm{Re}(\lambda_l)<0$，$l=1, 2, \cdots, n$。

一个矩阵 $A\in\mathbf{C}^{m\times m}$，如果它的所有特征值都有负实部，则称$A$是稳定矩阵。因而上面的一个定理就是说：$\frac{dx}{dt}=Ax$ 的零解是渐近稳定的，当且仅当A是稳定矩阵。

类似的方法可以证明

定理4.6.6 若A为常数矩阵，则$\frac{dx}{dt}=Ax$的零解在Lyapunov意义下是稳定的，当且仅当A的所有特征值都有非正的实部，并且实部为零的特征值所对应的Jordan块是一阶的。

证明 据定理4.6.3，只要证明上述充分必要条件与“e^{tA}是一致有界”等价就够了。再仿照定理 4.6.8 的证明，很容易看出e^{tA}是一致有界的，当且仅当 A 的特征值要么有负实部，要么有零实部，但对应的Jordan块是一阶的。

应用上述的两个定理来判定稳定性，要么计算出A的所有特征值，要么确定A的所有特征值在复平面上的位置。前者决非一件容易的事，所以不甚可行。后者关于特征值定位的方法，将在第五章给出。现在给出另外一种判定方法，它也是 Lyapunov 证明的。为此，先证明一个引理。

引理4.6.7 设 A，$V\in\mathbf{C}^{n\times n}$，$c\in\mathbf{C}^n$，$t\in\mathbf{R}$，并令$x(t)=e^{tA}c$，$f(t)=x^H V x$。则有

$$\frac{df}{dt}=x^H(A^H V+VA)x$$

证明 首先有$\frac{dx}{dt}=Ae^{tA}c=e^{tA}Ac$，$\frac{dx^H}{dt}=\left(\frac{dx}{dt}\right)^H=$

$c^H e^{tA^H} A^H$，从而有

$$\frac{df}{dt}=\frac{dx^H}{dt}Vx+x^HV\frac{dx}{dt}$$

$$=c^He^{tA^H}A^HVe^{tA}c+c^He^{tA^H}VAe^{tA}c$$

$$=c^He^{tA^H}(A^HV+VA)e^{tA}c$$

$$=x^H(A^HV+VA)x$$

定理4.6.8（Lyapunov定理） 矩阵$A\in\mathbf{R}^{n\times n}$为稳定矩阵，当且仅当存在唯一的实正定对称阵$V$满足方程

$$A^TV+VA=-I$$

证明 必要性：设A是实稳定矩阵，作线性映射F如下：

$$F:\ \mathbf{R}^{n\times n}\longrightarrow\mathbf{R}^{n\times n}$$

$$M\longmapsto A^TM+MA$$

即 $F(M)=A^TM+MA$。可以断言，F是一个可逆的映射，即F^{-1}存在。为此，我们证明F的零空间 $N(F)=\{0\}$。任取 $Z\in N(F)$，则有

$$F(Z)=A^TZ+ZA=0$$

即有$-A^TZ=ZA$，由此很容易看出，对任意的自然数k及$t\in\mathbf{R}$，总有$(-tA^T)^kZ=Z(tA)^k$，从而利用矩阵指数函数的定义可得到

$$e^{-tA^T}Z=Ze^{tA}$$

$$Z=e^{tA^T}Ze^{tA}$$

由于A^T，A均为稳定矩阵，故有

$$\lim_{t\to+\infty}e^{tA^T}=\lim_{t\to+\infty}e^{tA}=0$$

从而推出$Z=0$。由于F^{-1}存在，若取$V=F^{-1}(-I)$，则有

$$F(V)=A^TV+VA=-I$$

这表明定理要求的V是存在的。其实也是唯一的，因为若$F(W)=A^TW+WA=-I$时，便有

$$W=F^{-1}(-I)=V$$

留下来的工作是要证明V是正定对称阵。因为

$$F(V^T)=A^TV^T+V^TA=(VA)^T+(A^TV)^T$$
$$=(VA+A^TV)^T=(-I)^T=-I$$

据前证唯一性知$V^T=V$。为了证明V是正定的，现采用反证的方法。设有某一$c\in\mathbf{R}^n$且$c\neq 0$使得$c^TVc\leqslant 0$，则取$x=e^{At}c$，并作二次型

$$f(t)=x^TVx$$

从而据引理4.6.7可得

$$\frac{df}{dt}=x^T(A^TV+VA)x=x^T(-I)x=-x^Tx<0$$

这表明$f(t)$对一切$t\in\mathbf{R}$都是严格单调下降的，故而有

$$f(1)<f(0)=c^TVc\leqslant 0$$

因此对一切$t\geqslant 1$，更有$f(t)\leqslant f(1)<0$，当然$\lim\limits_{t\to+\infty}f(t)\neq 0$，但据$A$是稳定矩阵应有

$$\lim_{t\to+\infty}e^{tA}=0$$

$$\lim_{t\to+\infty}f(t)=\lim_{t\to+\infty}c^Te^{tA^T}Ve^{tA}=0$$

从而出现了矛盾。这表明V是正定的。

充分性：设矩阵方程$A^TV+VA=-I$有唯一的实正定对称阵V，并令λ，v为A的特征値及相应的特征向量，即$Av=\lambda v$。作二次型

$$f(t)=(e^{\lambda t}v)^HVe^{\lambda t}v$$

上式两端关于t求导可得

$$\frac{df}{dt}=\bar{\lambda}(e^{\lambda t}v)^HVe^{\lambda t}v+\lambda(e^{\lambda t}v)^HVe^{\lambda t}v$$
$$=(\bar{\lambda}+\lambda)f(t)=f(t)\cdot 2\mathrm{Re}(\lambda)$$

据$Av=\lambda v$可知，对任意的n都有$(tA)^nv=(t\lambda)^nv$，从而有

$$e^{tA}v=\sum_{n=0}^{\infty}\frac{(tA)^n}{n!}v=\sum_{n=0}^{\infty}\frac{(t\lambda)^n}{n!}v=e^{\lambda t}v$$

$$f(t)=(e^{tA}v)^H V e^{tA}v$$

$$\frac{df}{dt}=-(e^{tA}v)^H e^{tA}v=-e^{(\bar{\lambda}+\lambda)t}v^H v$$

$$=-e^{2t\mathrm{Re}(\lambda)}v^H v$$

与前面关于$\frac{df}{dt}$的推导结果合在一起便有

$$2f(t)\mathrm{Re}(\lambda)=-e^{2t\mathrm{Re}(\lambda)}v^H v$$

亦即

$$2(e^{\lambda t}v)^H V e^{\lambda t}v\mathrm{Re}(\lambda)=-e^{-2t\mathrm{Re}(\lambda)}v^H v$$

令$t=0$代入，并注意到V的正定性，使得

$$\mathrm{Re}(\lambda)<0$$

由于λ是A的任一特征值，故A为稳定矩阵。

上述的 Lyapunov 定理把判定A是否为稳定矩阵的问题转化到判定矩阵方程 $A^T V+VA=-I$ 有无唯一的对称正定解V的问题。所以在矩阵理论中，又把形如

$$AX+XB=C$$

的矩阵方程叫做 Lyapunov 方程。关于这个方程有很多研究论文，这里不加证明地给出一个结果：若对给定的A，B，积分$\int_0^{\infty}e^{tA}C e^{tB}dt$对所有的$C$都存在，则Lyapunov 方程$AX+XB=C$有唯一的解

$$X=-\int_0^{\infty}e^{tA}Ce^{tB}dt$$

方程$AX+XB=C$对一切C都有解的充分必要条件是$\lambda_i+\mu_j\neq 0$，其中λ_i是A的特征值，μ_j是B的特征值。在第二章我们曾用矩阵的 Kronecker 积证明过这个结论。

第七节　矩阵函数e^{tA}与Laplace变换

由于矩阵函数e^{tA}在线性系统理论及线性算子半群理论中有

着特殊的作用，所以这里还要对它作单独的讨论，主要是考虑它的一种计算方法，即 Laplace 变换，这与用 Laplace 变换求解线性常系数微分方程的思想是一脉相承。

在上节，我们证明定理4.6.5时曾推出

$$e^{tA}=P\left[\bigoplus_{i=1}^{k}e^{\lambda_i t}\sum_{j=0}^{n_i-1}\frac{(tU_{n_i})^j}{j!}\right]P^{-1}$$

显然，要想通过该式计算e^{tA}，先要知道A的Jordan 标准形及变换阵P，这是很困难的。如果我们把块对角阵

$$\bigoplus_{i=1}^{k}e^{\lambda_i t}\sum_{j=0}^{n_i-1}\frac{(tU_{n_i})^j}{j!}$$

$$=\operatorname{diag}\left(e^{\lambda_1 t}\sum_{j=0}^{n_1-1}\frac{(tU_{n_1})^j}{j!},\ e^{\lambda_2 t}\sum_{j=0}^{n_2-1}\frac{(tU_{n_2})^j}{j!},\right.$$

$$\left.\cdots,\ e^{\lambda_k t}\sum_{j=0}^{n_k-1}\frac{(tU_{n_k})^j}{j!}\right)$$

中的每一小块改写为

$$e^{\lambda_i t}\sum_{j=0}^{n_i-1}\frac{(tU_{n_i})^j}{j!}=\sum_{j=0}^{n_i-1}\left(\frac{t^j}{j!}e^{\lambda_i t}\right)U_{n_i}^j,$$

$$i=1,\ 2,\ \cdots,\ k$$

并令

$$A_{1j}=P\operatorname{diag}(U_{n_1}^j,\ O_{n_2},\ \cdots,\ O_{n_k})P^{-1}$$

$$j=0,\ 1,\ \cdots,\ n_1-1,$$

$$A_{2j}=P\operatorname{diag}(O_{n_1},\ U_{n_2}^j,\ \cdots,\ O_{n_k})P^{-1}$$

$$j=0,\ 1,\ \cdots,\ n_2-1$$

$$\vdots$$

$$A_{kj}=P\operatorname{diag}(O_{n_1},\ O_{n_2},\ \cdots,\ U_{n_k}^j)P^{-1},$$

$$j=0,\ 1,\ \cdots,\ n_k-1,$$

则有

$$e^{tA}=\sum_{i=1}^{k}\left[\sum_{j=0}^{n_i-1}\frac{t^j}{j!}e^{\lambda_i t}A_{ij}\right]=\sum_{i=1}^{k}e^{\lambda_i t}\sum_{j=0}^{n_i-1}\frac{t^j}{j!}A_{ij}$$

例如，若$A=P\text{diag}(J_3(2),\ J_2(3))P^{-1}$，则便有

$$e^{tA}=e^{2t}P\text{diag}(I_3,\ O_2)P^{-1}+te^{2t}P\text{diag}(U_3,\ O_2)P^{-1}$$
$$+\frac{t^2}{2}e^{2t}P\text{diag}(U_3^2,\ O_2)P^{-1}+$$
$$+e^{3t}P\text{diag}(O_3,\ I_2)P^{-1}+te^{3t}P\text{diag}(O_3,\ U_2)P^{-1}$$
$$=e^{2t}A_{10}+te^{2t}A_{11}+\frac{t^2}{2!}e^{2t}A_{12}+e^{3t}A_{20}+$$
$$+te^{3t}A_{21}$$
$$=\sum_{i=1}^{2}\sum_{j=1}^{n_i-1}\frac{t^j}{j!}e^{\lambda_i t}A_{ij}\quad(\lambda_1=2,\lambda_2=3,n_1=3,n_2=2)。$$

如果我们取λ_1，λ_2，…，λ_ν为A的所有互异特征值，并取ω足够大（$\omega>\max_i(n_i)$），则e^{tA}还可进一步写为

$$e^{tA}=\sum_{i=1}^{\nu}e^{\lambda_i t}\left(\sum_{j=0}^{\omega-1}\frac{t^j}{j!}B_{ij}\right)$$

其中B_{ij}是与A同阶的常数矩阵。例如前面的例子中的e^{tA}就可表为

$$e^{tA}=e^{2t}\left(A_{10}+tA_{11}+\frac{t^2A_{12}}{2!}+\frac{t^3O_5}{3!}+\frac{t^4O_5}{4!}\right)$$
$$+e^{2t}\left(A_{20}+tA_{21}+\frac{t^2O_5}{2!}+\frac{t^3O_5}{3!}+\frac{t^4O_5}{4!}\right)$$

这样表示，我们可以借助求解多项式

$$P_i(t)=\sum_{j=0}^{\omega-1}\frac{t^j}{z!}B_{ij}$$

的方法来计算e^{tA}无须先求A的Jordan 标准形及变换阵P。这就是下面要介绍的Laplace变换。

定义4.7.1（Laplace变换） 设$s\in\mathbf{C}$，$X(t)$为$n\times n$的方阵，若积分

$$L(s)=\int_0^{\infty}X(t)e^{-st}dt$$

存在，则把$L(s)$叫做$X(t)$的Laplace变换，并记作$\mathscr{L}(X(t))$。

根据矩阵函数的积分的定义，显然有

$$\mathscr{L}(X(t))=[\mathscr{L}(X_{ij}(t))]$$

定理4.7.2 若常数矩阵 A 的所有特征值 λ 都满足 $\mathrm{Re}(\lambda)<\mathrm{Re}(s)$，则有

$$\mathscr{L}(e^{tA})=(sI-A)^{-1}$$

证明 根据广义积分的定义，应有

$$\mathscr{L}(e^{tA})=\lim_{\tau\to\infty}\int_0^\tau e^{tA}e^{-st}dt$$

$$=\lim_{\tau\to\infty}\int_0^\tau e^{t(A-sI)}dt$$

由于 $\mathrm{Re}(\lambda)<\mathrm{Re}(s)$，故 s 不是 A 的特征值，从而 $(A-sI)$ 是可逆的。且因 $(A-sI)$ 的特征值有负实部，故 $(A-sI)$ 是稳定矩阵。从而有 $\lim\limits_{t\to\infty}e^{(A-sI)t}=0$。将这一结果用到 $\mathscr{L}(e^{tA})$ 中，便有

$$\mathscr{L}(e^{tA})=\lim_{\tau\to\infty}\int_0^\tau e^{t(A-sI)}dt$$

$$=\lim_{\tau\to\infty}(A-sI)^{-1}e^{t(A-sI)}\Big|_0^\tau$$

$$=\lim_{\tau\to\infty}(A-sI)^{-1}[e^{\tau(A-sI)}-I]$$

$$=(sI-A)^{-1}$$

根据 Laplace 变换的定义，很容易导出下列性质。

定理4.7.3 设标量函数 $f(t)$ 及矩阵函数 $X(t)$，$Y(t)$ 的 Laplace变换都存在，A是常数矩阵，则

1. $\mathscr{L}(f(t)A)=\mathscr{L}(f(t))A$
2. $\mathscr{L}(AX(t))=A\mathscr{L}(X(t))$
3. $\mathscr{L}(X(t)+Y(t))=\mathscr{L}(X(t))+\mathscr{L}(Y(t))$

因为标量函数 $t^n e^{\lambda t}$ 的Laplace变换为

$$\mathscr{L}(t^n e^{\lambda t})=\frac{n!}{(s-\lambda)^{n+1}},\quad n=0,1,2,\cdots,$$

其中Re(λ)<Re(s)。又因为矩阵函数e^{tA}可表为

$$e^{tA}=\sum_{i=1}^{\nu}e^{\lambda_i t}\sum_{j=0}^{\omega-1}\frac{t^j}{j!}B_{ij}$$

故据定理4.7.3中的性质，可得到

$$\begin{aligned}\mathscr{L}(e^{tA})&=\sum_{i=1}^{\nu}\mathscr{L}\left(e^{\lambda_i t}\sum_{j=0}^{\omega-1}\frac{t^j}{j!}\right)B_{ij}\\&=\sum_{i=1}^{\nu}\sum_{j=0}^{\omega-1}\mathscr{L}\left(\frac{t^j}{j!}e^{\lambda_i t}\right)B_{ij}\\&=\sum_{i=1}^{\nu}\sum_{j=0}^{\omega-1}\frac{1}{(s-\lambda_i)^{j+1}}B_{ij}\end{aligned}$$

从而得到等式

$$(sI-A)^{-1}=\sum_{i=1}^{\nu}\sum_{j=0}^{\omega-1}\frac{1}{(s-\lambda_i)^{j+1}}B_{ij}$$

我们的目的就是想通过这个等式把诸 B_{ij} 确定下来，以便得到e^{tA}的显然表达式。为此，要利用$(sI-A)$的伴随矩阵adj$(sI-A)$及行列式$|sI-A|$来表示$(sI-A)^{-1}$，即

$$(sI-A)^{-1}=\frac{1}{\det(sI-A)}\mathrm{adj}(sI-A)$$

兹举一例，以示说明。

例 若$A=\begin{bmatrix}1 & 4\\3 & 2\end{bmatrix}$，则 $\det(sI-A)=(s+2)(s-5)$，故$e^{tA}=e^{5t}C+e^{-2t}D$，其中C，D为常数阵。另一方面，又有(对上式两端取Laplace变换)：

$$\begin{aligned}(sI-A)^{-1}&=\frac{1}{(s+2)(s-5)}\begin{bmatrix}s-2 & 4\\3 & s-1\end{bmatrix}\\&=\frac{1}{s-5}C+\frac{1}{s+2}D\end{aligned}$$

因此便得到

$$(s+2)C+(s-5)D=\begin{bmatrix}s-2 & 4\\3 & s-1\end{bmatrix}$$

分别令$s=-2$，$s=5$，代入上式两端，便得到

$$C=\frac{1}{7}\begin{bmatrix}3 & 4\\ 3 & 4\end{bmatrix},\quad D=\frac{1}{7}\begin{bmatrix}4 & -4\\ -3 & 3\end{bmatrix}$$

从而得到

$$e^{tA}=\frac{1}{7}e^{5t}\begin{bmatrix}3 & 4\\ 3 & 4\end{bmatrix}+\frac{1}{7}e^{-2t}\begin{bmatrix}4 & -4\\ -3 & 3\end{bmatrix}$$

算例表明，用 Laplace 变换的方法计算e^{tA}也不是很容易的。只是对于某些特殊的A，譬如A的阶次很低，或A有n个不同的特征值，或A的n个特征值都相同。一般情况下，计算也是相当繁琐的。首先，我们假定$A\in\mathbf{C}^{n\times n}$有$n$个互异的特征值，这时$e^{tA}=\sum_{i=1}^{\nu}\sum_{j=0}^{\omega-1}e^{\lambda_i t}\frac{t^j}{j!}B_{ij}$可写为

$$e^{tA}=\sum_{i=1}^{n}e^{\lambda_i t}A_i$$

对该式两端取Laplace变换，便得到

$$(sI-A)^{-1}=\sum_{i=1}^{n}(s-\lambda_i)^{-1}A_i$$

而因$\det(sI-A)=\prod_{i=1}^{n}(s-\lambda_i)$，故又有

$$(sI-A)^{-1}=\frac{1}{\det(sI-A)}\operatorname{adj}(sI-A)$$

从而得到

$$\operatorname{adj}(sI-A)=\sum_{i=1}^{n}\prod_{l\neq i}(s-\lambda_l)A_i$$

分别令$s=\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$，代入，可得到

$$\operatorname{adj}(\lambda_j I-A)=\prod_{l\neq j}(\lambda_j-\lambda_l)A_j,\quad j=1,\ \cdots,\ n$$

$$A_j=\Big[\prod_{l\neq j}(\lambda_j-\lambda_l)\Big]^{-1}\operatorname{adj}(\lambda_j I-A),\quad j=1,\ 2,\ \cdots,\ n$$

从而得到

$$e^{tA}=\sum_{j=1}^{n}e^{\lambda_j t}A_j=\sum_{j=1}^{n}e^{\lambda_j t}\prod_{l\neq j}(\lambda_j-\lambda_l)^{-1}\mathrm{adj}(\lambda_j I-A)$$

该式表明，欲求 e^{tA}，需求A的所有特征值λ_j，及 n 个伴随矩阵 $\mathrm{adj}(\lambda_j I-A)$。

关于A有 n 个相同特征值的情形，先分析 $A=\begin{bmatrix}0 & -1\\ 4 & 4\end{bmatrix}$ 的简单情况。这时$\det(sI-A)=(s-2)^2$，故 $e^{tA}=e^{2t}(C+tD)$，两端取Laplace变换得

$$(sI-A)^{-1}=\frac{1}{s-2}C+\frac{1}{(s-2)^2}D$$

$$(sI-A)^{-1}=\frac{1}{\det(sI-A)}\mathrm{adj}(sI-A)$$

$$=\frac{1}{(s-2)^2}\begin{bmatrix}s-4 & -1\\ 4 & s\end{bmatrix}$$

从而有

$$\frac{1}{(s-2)^2}\begin{bmatrix}s-4 & -1\\ 4 & s\end{bmatrix}=\frac{1}{s-2}C+\frac{1}{(s-2)^2}D$$

即
$$\begin{bmatrix}s-4 & -1\\ 4 & s\end{bmatrix}=(s-2)C+D$$

两端关于 s 求导又得

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=C$$

前一等式中取$s=2$，便得

$$\begin{bmatrix}-2 & -1\\ 4 & 2\end{bmatrix}=D$$

最后便得到

$$e^{tA}=e^{2t}I+te^{2t}\begin{bmatrix}-2 & -1\\ 4 & 2\end{bmatrix}$$

将上述结果推广到 A 为高阶的情况，设 A 的特征多项式为 $(\lambda-\lambda_0)^n$，则e^{tA}的表达式应为

$$e^{tA}=\sum_{j=0}^{n-1}e^{\lambda_0 t}\frac{t^j}{j!}B_j$$

两端取Laplace变换可得

$$(sI-A)^{-1}=\sum_{j=1}^{n}(s-\lambda_0)^{-j}B_{j-1}$$

为下面的记法清晰起见，令 $\mathrm{adj}(sI-A)=B(s)$，则又有

$$(sI-A)^{-1}=\frac{1}{\det(sI-A)}\mathrm{adj}(sI-A)=(s-\lambda_0)^{-n}B(s)$$

比较后可知

$$B(s)=\sum_{j=1}^{n}(s-\lambda_0)^{n-j}B_{j-1}$$

两端关于 s 求导 i 次，便得

$$B^{(i)}(s)=\sum_{j=1}^{n-i}(n-j)(n-j-1)\cdots(n-j-i+1)(s-\lambda_0)^{n-j-i}B_{j-1}$$

再令$s=\lambda_0$ 代入，便得到

$$B^{(i)}(\lambda_0)=i!B_{n-i-1},\quad 0\leqslant i\leqslant n-1$$

如此便解得

$$B_j=\frac{1}{(n-j-1)!}B^{(n-j-1)}(\lambda_0),\quad 0\leqslant j\leqslant n-1$$

最后得到

$$e^{tA}=e^{\lambda_0 t}\sum_{j=0}^{n-1}\frac{t^j}{(n-j-1)!j!}B^{(n-j-1)}(\lambda_0)$$

综上所述，可得如下结论：

定理4.7.4 设$A\in \mathbf{C}^{n\times n}$，则有

1. 若A有 n 个互异的特征值λ_1，λ_2，$\cdots\lambda_n$，则

$$e^{tA}=\sum_{j=1}^{n}e^{\lambda_j t}\prod_{l\neq j}(\lambda_j-\lambda_l)^{-1}\mathrm{adj}(\lambda_j I-A)$$

2. 若A只有一个n重特征值λ_0，且记$B^{(j)}(s)$为$\mathrm{adj}(sI-A)$关于s的第j次导数，则

$$e^{tA}=e^{\lambda_0 t}\sum_{j=0}^{n-1}\frac{t^j}{(n-j-1)!\,j!}B^{(n-j-1)}(\lambda_0)$$

对于一般的A，虽然按照上述方法也可以计算出e^{tA}，但是却很难象上面那样给出一个通式，这里用一个算例说明计算的过程。

例1 设$A\in\mathbf{C}^{6\times 6}$的特征多项式为

$$\det(\lambda I-A)=(\lambda-1)^3(\lambda-3)^2(\lambda+3)$$

则e^{tA}有展开式

$$e^{tA}=e^t\sum_{j=0}^{2}\frac{t^j}{j!}C_j+e^{3t}\sum_{j=0}^{1}\frac{t^j}{j!}D_j+e^{-3t}E$$

等号两端取Laplace变换便得

$$(sI-A)^{-1}=\frac{1}{s-1}C_0+\frac{1}{(s-1)^2}C_1+\frac{1}{(s-1)^3}C_2$$

$$+\frac{1}{s-3}D_0+\frac{1}{(s-3)^2}D_1+\frac{1}{s+3}E$$

又据$(sI-A)^{-1}=[\det(sI-A)]^{-1}\mathrm{adj}(sI-A)$可得

$$\begin{aligned}B(s)&\triangleq\mathrm{adj}(sI-A)\\&=(s-3)^2(s+3)[(s-1)^2C_0+(s-1)C_1+C_2]\\&\quad+(s-1)^3(s+3)[(s-3)D_0+D_1]+\\&\quad(s-1)^3(s-3)^2E\end{aligned}$$

由此可得

$$B(1)=16C_2,\qquad B(3)=48D_1,\qquad B(-3)=-2304E$$

故有 $C_2=\frac{1}{16}B(1)$，$D_1=\frac{1}{48}B(3)$，$E=-\frac{1}{2304}B(-3)$

对$B(s)$求导又得

$$B'(s)=[2(s-3)(s+3)+(s-3)^2][(s-1)^2C_0$$
$$+(s-1)C_1+C_2]+$$

$$(s-1)^3(s+3)[2(s-1)C_0+C_1]+$$
$$[3(s-1)^2(s+3)+(s-1)^3][(s-3)D_0+D_1]+$$
$$(s-1)^3(s+3)D_0+[3(s-1)^2(s-3)^2+$$
$$2(s-1)^3(s-3)]E$$

由此可得

$$B'(1)=-12C_2+16C_1=-\frac{12}{16}B(1)+16C_1$$

$$B'(3)=80D_1+48D_0=\frac{80}{48}B(3)+48D_0$$

这样便解出

$$C_1=\frac{3}{64}B(1)+\frac{1}{16}B'(1),$$

$$D_0=-\frac{5}{144}B(3)+\frac{1}{48}B'(3)$$

对$B'(s)$再求导一次，又得

$$B''(s)=[4s+2(s-3)][(s-1)^2C_0+(s-1)C_1+C_2]+$$
$$[2s^2+18-(s-1)^2][2(s-1)C_0+C_1]+$$
$$[2(s-3)(s+3)+(s-3)^2][2(s-1)C_0+$$
$$C_1]+(s-1)P(s)$$

其中$P(s)$是以矩阵为系数的s的某一多项式，因为$B'(s)$的后三项都含有因子$(s-1)^2$，故求导后能提出公因子$(s-1)$。从而令$s=1$代入便有

$$B''(1)=-24C_1+32C_0$$
$$=-\frac{9}{8}B(1)-\frac{3}{2}B'(1)+32C_0$$

由此解得

$$C_0=\frac{9}{256}B(1)+\frac{3}{64}B'(1)+\frac{1}{32}B''(1)$$

最后，将C_0，C_1，C_2，D_0，D_1，E代入便求得

$$e^{tA}=e^{t}\Big[\Big(\frac{9}{256}B(1)+\frac{3}{64}B'(1)+\frac{1}{32}B''(1)\Big)+$$

$$t\Big(\frac{3}{64}B(1)+\frac{1}{16}B'(1)\Big)+\frac{t^2}{32}B(1)\Big]+$$

$$e^{3t}\Big[\Big(-\frac{5}{144}B(3)+\frac{1}{48}B'(3)\Big)+\frac{t}{48}B(3)\Big]-$$

$$\frac{1}{2304}B(-3)。$$

从上面的算例可以看出，每个C_j都是$B(1),B'(1)$，$B''(1)$的线性组合，而每个D_j也都是$B(3)$，$B'(3)$的线性组合，E是$B(-3)$的一个倍数。这一现象并非偶然的巧合。

通常的情况下，A的特征多项式如果是

$$\prod_{j=1}^{\nu}(\lambda-\lambda_j)^{n_j}$$

其中λ_1，$\lambda_2,\cdots,\lambda_\nu$是$A$的互异特征值，$n_j$为$\lambda_j$的重数。则必有

$$e^{tA}=\sum_{i=1}^{\nu}e^{\lambda_i t}\sum_{j=0}^{n_i-1}\frac{t^j}{j!}B_{ij}$$

其中B_{ij}一定是$B(s)=\text{adj}(sI-A)$及其各阶导数在λ_i处值的线性组合，亦即

$$B_{ij}=\sum_{l=0}^{n_i-1}c_l B^{(l)}(\lambda_i),\qquad i=1,2,\cdots,\nu$$

我们看到，用 Laplace 变换的方法计算e^{tA}虽然避开了将A化为 Jordan 标准形的困难，但却要付出计算（$sI-A$）的伴随矩阵$\text{adj}(sI-A)$及其各阶导数的代价。当A阶数很高时，其复杂的程度是难以想象的。这里倒有一种避开计算$(sI-A)_{ij}$的代数余子式以求$\text{adj}(sI-A)$的方法，它是一种递推的算法。仍记$B(s)=\text{adj}(sI-A)$，而把$\det(sI-A)$记作为$f(s)=s^n+a_1s^{n-1}+\cdots+a_n$，则有关于$s$的恒等式

$$(sI-A)B(s)=f(s)I$$

对上式两端反复求异便得到：

$$B(s)=(A-sI)B'(s)+f'(s)I$$

$$2B'(s)=(A-sI)B''(s)+f''(s)I$$

$$\vdots$$

$$(k-1)B^{(k-2)}(s)=(A-sI)B^{(k-1)}(s)+f^{(k-1)}(s)I$$

$$kB^{(k-1)}(s)=(A-sI)B^{(k)}(s)+f^{(k)}(s)I$$

$$\vdots$$

$$nB^{(n-1)}(s)=(A-sI)B^{(n)}(s)+f^{(n)}(s)I$$

注意到$B(s)$的每个元素皆是次数低于 n 的 s 的多项式，故，$B^{(n)}(s)=0$,而 $f^{(n)}(s)=n!$，故有

$$B^{(n-1)}(s)=(n-1)!I$$

$$B^{(n-2)}(s)=[(A-sI)B^{(n-1)}(s)+f^{(n-1)}(s)I]/(n-1)$$

$$B^{(n-3)}(s)=[(A-sI)B^{(n-2)}(s)+f^{(n-2)}(s)I]/(n-2)$$

$$\vdots$$

$$B'(s)=[(A-sI)B''(s)+f''(s)I]/2$$

$$B(s)=(A-sI)B'(s)+f'(s)I$$

这样就能递推地算出 $B^{(n-1)}(s)$，$B^{(n-2)}(s)$，…，$B'(s)$，$B(s)$，从而再仿照例 1 算出e^{tA}。不过A 的特征多项式和特征值还是要算出来的。

例2 若取

$$A=\begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$$

则其特征多项式为$f(s)=(s-5)(s-1)^2$，据此求得

$$f'(s)=3s^2-14s+11$$

$$f''(s)=6s-14$$

$$B''(s)=2I$$

$$B'(s)=\frac{1}{2}[A-sI)(2I)+(6s-14)I]=A+(2s-7)I$$

$$B(s)=(A-sI)[A+(2s)-7)I]+(3s^2-14s+11)I$$
$$=A^2+(s-7)A+(s^2-7s+11)I$$

第八节 例题与习题

1． 证明：

$$\| A \|_\infty = \sup_{\|x\|_\infty=1} \| Ax \|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

$$\| A \|_2 = \sup_{\|x\|_2=1} \| Ax \|_2 = \max\{\sigma_1, \sigma_2, \cdots\sigma_n\}$$

其中σ_i为A的奇异值。

2、 设 $\|\cdot\|$ 为 $\mathbf{C}^m$ 上的一种向量范数，对固定的矩阵 $A\in\mathbf{C}^{m\times n}$，函数 $f(x)=\|Ax\|$ 能作为$\mathbf{C}^n$上的范数，其充分必要条件是$\text{rank}(A)=n$。

提示：必要性。用反证法，若$\|Ax\|$是$\mathbf{C}^n$中x的范数，但$\text{rank}(A)<n$。则存在$0\neq x_0\in\mathbf{C}^n$，使得$Ax_0=0$，从而有

$$\| Ax_0 \| =0$$

这说明$\mathbf{C}^n$中非零向量x_0的范数可以为零，与范数定义矛盾。

充分性。令 $A=(a_1, a_2, \cdots, a_n)$，因 $\text{rank}(A)=n$，故$\{a_1, a_2, \cdots, a_n\}$是线性无关的。所以对一切非零的向量 $x\in\mathbf{C}^n$有$Ax\neq0$，从而$\|Ax\|\neq0$。对任意的$\alpha\in\mathbf{C}$，显然有

$$\| A(\alpha x)\| = \|\alpha Ax\| = |\alpha|\,\|Ax\|$$

此外，对一切x，$y\in\mathbf{C}^n$有

$$\begin{aligned}\| A(x+y)\| &= \| Ax+Ay\| \\ &\leqslant \|Ax\| + \|Ay\|\end{aligned}$$

所以满足范数的三条公理。

3． 设$\{\alpha_1, \alpha_2, \cdots, \alpha_n\}$是一组给定的正数，$x\in\mathbf{C}^n$且$x=(\xi_1, \xi_2, \cdots, \xi_n)^T$，则函数

$$f(x)=\left[\sum_{i=1}^n \alpha_i |\xi_i|^2\right]^{1/2}$$

在$\mathbf{C}^n$上定义了一种范数，试证明之。

4． 证明

$$\|AB\|_F \leqslant \min\{\|A\|_2 \cdot \|B\|_F,\ \|A\|_F \cdot \|B\|_2\}$$

5． 设$A \in \mathbf{C}^{m\times n}$是部分酉阵，即$A^H A = I_n$。试证

$$\|A\|_2 = 1,\ \|A\|_F = \sqrt{n}$$

6． 证明：对任一种算子范数$\|\cdot\|$，都有

$$\|I\| = 1,\ \|A^{-1}\| \geqslant \|A\|^{-1}$$

7． 设$A \in \mathbf{C}_n^{n\times n}$，$\lambda$为$A$的特征值。试证：

$$\|A^{-1}\|_2^{-1} \leqslant |\lambda| \leqslant \|A\|_2$$

在数值分析中遇到这样的问题：给予可逆矩阵A一个摄动δA，试问

（a） 在什么条件下，$A+\delta A$仍可逆？

（b） 如果$A+\delta A$可逆，那么$(A+\delta A)^{-1}$在多大的程度上逼近A^{-1}？下面的两个习题是一个回答。

8． 若$\|A\|<1$，则$I-A$可逆且

（a） $\|(I-A)^{-1}\| \leqslant \dfrac{1}{1-\|A\|}$

（b） $\|(I-A)^{-1}-I\| \leqslant \dfrac{\|A\|}{1-\|A\|}$

提示：任取$x \in \mathbf{C}^n$都有

$$\|(I-A)x\| \geqslant \|x\| - \|Ax\| \geqslant (1-\|A\|)x$$

这表明，只要$x \neq 0$，$(I-A)x$也不为零，从而说明$I-A$是可逆的。于是有

$$(I-A)(I-A)^{-1} = I$$

上式左端展开并移项可得

$$(I-A)^{-1} = I + A(I-A)^{-1} \qquad (*)$$

取范数，再利用三角不等式便得

$$\|(I-A)^{-1}\| \leqslant 1 + \|A\|\,\|(I-A)^{-1}\|$$

整理后便证明了(a)。利用($*$)及(a)，又有

$$\|(I-A)^{-1}-I\|=\|A(I-A)^{-1}\|$$
$$\leqslant\|A\|\,\|(I-A)^{-1}\|$$
$$\leqslant\frac{\|A\|}{1-\|A\|}$$

从而证明了(b)。

9. 设 A 是可逆的，且 $\|A^{-1}\delta A\|<1$。则 $A+\delta A$ 亦可逆，且有

(a) $(A+\delta A)^{-1}=(I+F)A^{-1}$

其中

$$\|F\|\leqslant\frac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$$

(b) $$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|}\leqslant\frac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$$

提示：(a) 由于A可逆，故 $A+\delta A=A[I+A^{-1}\delta A]$，又因 $\|A^{-1}\delta A\|<1$，据习题8，$I+A^{-1}\delta A$ 可逆。从而 $A+\delta A$ 可逆，且有

$$(A+\delta A)^{-1}=(I+A^{-1}\delta A)^{-1}A^{-1}$$

若令 $(I+A^{-1}\delta A)^{-1}=I+F$，则上式为

$$(A+\delta A)^{-1}=(I+F)A^{-1}\qquad(**)$$

又据 $F=(I+A^{-1}\delta A)^{-1}-I$ 可得

$$\|F\|=\|I-(I+A^{-1}\delta A)^{-1}\|$$
$$\leqslant\frac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$$

(b)据$(**)$有 $A^{-1}-(A+\delta A)^{-1}=-FA^{-1}$，取范数便得

$$\|A^{-1}-(A+\delta A)^{-1}\|\leqslant\|F\|\,\|A^{-1}\|$$

再据(a)便证明了(b)。

10. 可逆矩阵$A\in\mathbf{C}^{n\times n}$的条件数定义为

$$k(A)=\|A\|\,\|A^{-1}\|$$

对于不同的算子范数，诸如$\|\cdot\|$，$\|\cdot\|_2$及$\|\cdot\|_F$等，其相应

的条件数$k_1(A)$，$k_2(A)$，及$k_F(A)$等一般是不同的。但不管是哪一种算子范数，总有

$$k(A)\geqslant 1$$

这是显然的。说明矩阵条件数的下界为1，請证明酉矩阵关于算子范数$\|\cdot\|_2$，其条件数是1。

11. 设$A\in \mathbf{C}_n^{n\times n}$，$\delta A\in \mathbf{C}^{n\times n}$，且$\|A^{-1}\|\|\delta A\|<1$。试证习题9中的$F$满足

$$\|F\|\leqslant \frac{k(A)\dfrac{\|\delta A\|}{\|A\|}}{1-k(A)\dfrac{\|\delta A\|}{\|A\|}}$$

而逆矩阵的相对误差则满足

$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|}\leqslant \frac{k(A)\dfrac{\|\delta A\|}{\|A\|}}{1-k(A)\dfrac{\|\delta A\|}{\|A\|}}$$

提示：只要利用函数 $\varphi(x)=\dfrac{x}{1-x}$在（0，1）上的单调递增性及习题9的结论就行了。这个结论表明，A的条件数$k(A)$刻画了矩阵A摄动前后其逆之间的相对误差界。当$k(A)$较大时，矩阵A便称之为病态的，否则，叫做良态的。

12. 如果要问：当$b\in \mathbf{C}^n$或$A\in \mathbf{C}_n^{n\times n}$产生一定的摄动时，会对方程$Ax=b$的解带来多大的影响呢？对此，有下述结论：

(a) 若$\tilde{x}$满足

$$A\tilde{x}=\tilde{b}$$

则有

$$\frac{\|x-\tilde{x}\|}{\|x\|}\leqslant k(A)\frac{\|b-\tilde{b}\|}{\|b\|}$$

(b) 若x^*满足

$$(A+\delta A)x^*=b$$

且$\|A^{-1}\delta A\|<1$，则有

$$\frac{\|x-x^*\|}{\|x\|}\leqslant\frac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$$

若进一步满足$\|A^{-1}\|\,\|\delta A\|<1$时，有

$$\frac{\|x-x^*\|}{\|x\|}\leqslant\frac{k(A)\frac{\|\delta A\|}{\|A\|}}{1-k(A)\frac{\|\delta A\|}{\|A\|}}$$

提示：(a)因为，$(x-\tilde{x})=A^{-1}(b-\tilde{b})$，故有

$$\|x-\tilde{x}\|\leqslant\|A^{-1}\|\cdot\|b-\tilde{b}\|$$

又因为

$$\|b\|=\|Ax\|\leqslant\|A\|\,\|x\|$$

两式结合起来，便有

$$\frac{\|x-\tilde{x}\|}{\|x\|}\leqslant\frac{\|A^{-1}\|\,\|b-\tilde{b}\|\,\|A\|}{\|b\|}$$

$$=k(A)\frac{\|b-\tilde{b}\|}{\|b\|}$$

(b)利用习题9的结论，有

$$x-x^*=[A^{-1}-(A+\delta A)^{-1}]b$$
$$=-FA^{-1}b=-Fx$$

再取范数，利用习题9及习题11，便证明了(b)。

13．构造一个2×2阶的可逆矩阵序列，其收敛于不可逆的极限。

14．证明：当且仅当$A=I$时有$\lim\limits_{n\to\infty}A^n=I$。

15．若$A_k\in\mathbf{C}^{m\times n}$且$\sum\limits_{k=0}^{\infty}A_k<\infty$，则$\lim\limits_{k\to\infty}A_k=0$。但它的逆命题不真，试举反例。

16．证明：对于$A\in\mathbf{C}^{n\times n}$有恒等式

$$\cos 2A=\cos^2A-\sin^2A,\quad \sin^2A=\frac{1}{2}(I-\cos 2A)$$

17．计算下面级数的和：

$$\sum_{n=0}^{\infty}\begin{pmatrix}0.1 & 0.7\\0.3 & 0.6\end{pmatrix}^n$$

18． 证明：若 $A\in\mathbf{C}^{m\times m}$，$\rho(A)<1$，则 $\sum_{n=0}^{\infty} nA^n = A(I-A)^{-2}$。

19． 计算：

$$\sum_{n=0}^{\infty}\frac{n}{10^n}\begin{bmatrix}1 & 2\\8 & 1\end{bmatrix}^n$$

20． 设 $A=\begin{pmatrix}0 & -1\\4 & 1\end{pmatrix}$，计算 e^A。

21． 设

$$A=\begin{pmatrix}-1 & 2 & -2 & 2\\-2 & 2 & -1 & 3\\-1 & 0 & 1 & 2\\-1 & 1 & -1 & 2\end{pmatrix}$$

计算 $\tan^{-1}\left(A^{30}-\frac{1}{2}A^{20}\right)$

22． 给定

$$A=\begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$$

计算 e^{tA}

提示：首先有

$$e^{tA}=I+tA+\frac{t^2}{2!}A^2+\cdots$$

又知 A 的特征方程为

$$f(\lambda)=\lambda^2+1=0$$

据Cayley-Hamilton定理知

$$f(A)=A^2+I=0$$

据此可得

$$A^2=-I,\ A^3=-A,\ A^4=I,\ A^5=A,\ \cdots$$

从而有

$$e^{tA}=I+tA-\frac{t^2}{2!}I-\frac{t^3}{3!}A+\frac{t^4}{4!}I+\frac{t^5}{5!}A-\cdots$$

$$=I\left(1-\frac{t^2}{2!}+\frac{t^4}{4!}-\cdots\right)+A\left(t-\frac{t^3}{3!}+\frac{t^5}{5!}-\cdots\right)$$

$$=I\cos t+A\sin t$$

可见

$$E^{t\begin{bmatrix}0&1\\-1&0\end{bmatrix}}=\begin{bmatrix}\cot t&\sin t\\-\sin t&\cos t\end{bmatrix}$$

23. 设

$$A=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\end{bmatrix}$$

计算e^{tA}。

提示：A的特征方程为$f(\lambda)=\lambda^3=0$，故据$C-H$ 定理有 $A_3=0$，从而对$r\geqslant3$，有$A^r=0$。这样，便得到

$$e^{tA}=I+tA+\frac{t^2}{2!}A^2=\begin{bmatrix}1&0&0\\t&1&0\\t^2/2&t&1\end{bmatrix}$$

24. 设

$$A=\begin{pmatrix}1&2\\2&1\end{pmatrix}$$

计算e^A。

提示：据第二章第七节习题19，

$$A^r=rA+(1-r)I$$

$$\frac{A^r}{r!}=\frac{A}{(r-1)!}+\frac{1-r}{r!}I$$

$$e^A=\sum_{r=0}^{\infty}\frac{A^r}{r!}=I+A\sum_{r=1}^{\infty}\frac{1}{(r-1)!}+I\sum_{r=1}^{\infty}\frac{1-r}{r!}$$

$$=I+eA+(e-1-e)I=eA$$

$$=\begin{pmatrix} e & 2e \\ 0 & e \end{pmatrix}$$

25． 设e^A是一个可对角化的矩阵，试证：

$$\det(e_A)=e^{tr(A)}$$

特别当A是斜对称阵时，有 $\det(e^A)=1$。

26． 证明：若A为实反对称阵，则e^A为正交阵。

27． 证明：若A为Hermite阵，则e^{iH} 为酉矩阵。

28． 证明：

$$e^{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}}=\begin{pmatrix} e & 0 \\ 0 & e^2 \end{pmatrix}$$

29． 证明：若$AB=BA$，则$e^A \cdot e^B=e^B \cdot e^A=e^{A+B}$。并用习题22的结果和本习题的结论计算

$$e^{\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}}$$

30． 用谱上一致多项式的概念计算矩阵函数：

(a) $A=\begin{pmatrix} 0 & 1 \\ 0 & -2 \end{pmatrix}$，求$e^{tA}$；

(b) $\begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{pmatrix}$计算$e^{tA}$，$\sin tA$

提示：(a)先求A的特征多项式

$$f(\lambda)=\begin{vmatrix} \lambda & -1 \\ 0 & \lambda+2 \end{vmatrix}=\lambda(\lambda+2)$$

由此得到特征值 $\lambda_1=0, \lambda_2=-2$。由特征值互异知$A$是可对角化的，故最小多项式为

$$m_A(\lambda)=f(\lambda)=\lambda^2+2\lambda$$

从而知与$g(\lambda)=e^{\lambda t}$谱上一致的多项式应为

$$a_0(t)+a_1(t)\lambda$$

由此得到方程组

$$\begin{cases}\alpha_0(t)+\alpha_1(t)\lambda_1=e^{\lambda_1 t}\\ \alpha_0(t)+\alpha_1(t)\lambda_2=e^{\lambda_2 t}\end{cases}$$

令$\lambda_1=0$，$\lambda_2=-2$代入，解得

$$\alpha_0(t)=1, \ \alpha_1(t)=-\frac{1}{2}(e^{-2t}-1)$$

最后便得到

$$\begin{aligned}e^{tA}&=\alpha_0(t)I+\alpha_1(t)A\\ &=I-\frac{1}{2}(e^{-2t}-1)A\end{aligned}$$

不过，也可以根据$A^2=-2A$推出

$A^3=(-2)^2A$，$A^4=(-2)^3A$，…，$A^n=(-2n)^{n-1}A$，

从而求得

$$\begin{aligned}e^{tA}&=\sum_{n=0}^{\infty}\frac{t^nA^n}{n!}=I+\sum_{n=1}^{\infty}\frac{t^n}{n!}A^n\\ &=I+\sum_{n=1}^{\infty}\frac{t^n}{n!}(-2)^{n-1}A\\ &=I-\frac{1}{2}A\sum_{n=1}^{\infty}\frac{(-2t)^n}{n!}\\ &=I-\frac{1}{2}(e^{-2t}-1)A\end{aligned}$$

（b）先求A的特征多项式

$$f(\lambda)=\begin{vmatrix}\lambda-2 & -1 & -4\\ 0 & \lambda-2 & 0\\ 0 & -3 & \lambda-1\end{vmatrix}=(\lambda-2)^2(\lambda-1)$$

易见特征值$\lambda_1=\lambda_1=2$，$\lambda_2=1$；通过验算可知$(A-2I)(A-I)\neq 0$，故$(\lambda-2)(\lambda-1)$不是A的零化多项式，故可断定最小多项式为

$$m_A(\lambda)=(\lambda-2)^2(\lambda-1)$$

由于最小多项式的次数为3，故谱上一致多项式可设为

$$\alpha_0(t)+\alpha_1(t)\lambda+\alpha_2(t)\lambda^2$$

由于λ_1是重根，故求导一次代入，仍能得到

$$\begin{cases} g(\lambda_1)=\alpha_0(t)+\alpha_1(t)\lambda_1+\alpha_2(t)\lambda_1^2 \\ g'(\lambda_1)= \qquad \alpha_1(t) \quad +2\alpha_2(t)\lambda_1 \\ g(\lambda_2)=\alpha_0(t)+\alpha_1(t)\lambda_2+\alpha_2(t)\lambda_2^2 \end{cases} \qquad (*)$$

令$g(\lambda)=e^{\lambda t}$，$\lambda_1=2$，$\lambda_2=1$代入上面方程组，便得到

$$\begin{cases} \alpha_0(t)+2\alpha_1(t)+4\alpha_2(t)=e^{2t} \\ \alpha_1(t)+4\alpha_2(t)=te^{2t} \\ \alpha_0(t)+\alpha_1(t)+\alpha_2(t)=e^t \end{cases}$$

解之得：

$$\alpha_0(t)=4e^t-3e^{2t}+2te^{2t}$$

$$\alpha_1(t)=-4e^t+4e^{2t}-3te^{2t}$$

$$\alpha_2(t)=e^t-e^{2t}+te^{2t}$$

将$\alpha_0(t)$，$\alpha_1(t)$，$\alpha_2(t)$代入

$$e^{tA}=\alpha_0(t)I+\alpha_1(t)A+\alpha_2(t)A^2$$

便得到所要求的结果。

再令$g(\lambda)=\sin\lambda t$，$\lambda_1=2$，$\lambda_2=1$代入（$*$），便得到

$$\begin{cases} \alpha_0(t)+2\alpha_1(t)+4\alpha_2(t)=\sin 2t \\ \alpha_1(t)+4\alpha_2(t)=t\cos 2t \\ \alpha_0(t)+\alpha_1(t)+\alpha_2(t)=\sin t \end{cases}$$

由此解出$\alpha_0(t)$，$\alpha_1(t)$，$\alpha_2(t)$，代入

$$\sin At=\alpha_0(t)I+\alpha_1(t)A+\alpha_2(t)A^2$$

便得到所欲求的结果。

31．设

$$A(t)=\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

求$\frac{d}{dt}(A(t))$，$\frac{d}{dt}(A^{-1}(t))$，$\frac{d}{dt}|A(t)|$，$\left|\frac{d}{dt}A(t)\right|$。

32. 证明：如果A的所有特征值λ都位于左半平面，即$\mathrm{Re}(\lambda)<0$，则$\lim\limits_{t\to\infty} e^{tA}=0$。

33. 证明定理4.7.3的结论。

34. 利用拉氏变换求出矩阵e^{At}的显式表示，其中

(a) $A=\begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$，

(b) $A=\begin{bmatrix} 1 & -4 & -1 & -4 \\ 2 & 0 & 5 & -4 \\ -1 & 1 & -2 & 3 \\ -1 & 4 & -1 & 6 \end{bmatrix}$

提示(a)A的特征多项式为

$$|\lambda I-A|=\begin{vmatrix} \lambda-2 & -2 & -1 \\ -1 & \lambda-3 & -1 \\ -1 & -2 & \lambda-2 \end{vmatrix}=\begin{vmatrix} \lambda-5 & -2 & -1 \\ \lambda-5 & \lambda-3 & -1 \\ \lambda-5 & -2 & \lambda-2 \end{vmatrix}$$

$$=(\lambda-5)\begin{vmatrix} 1 & -2 & -1 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & \lambda-1 \end{vmatrix}=(\lambda-1)^2(\lambda-5)$$

如果$A\in \mathbf{C}^{n\times n}$有$k$个不同的特征值$\lambda_1$，…，$\lambda_k$，其代数重数分别为$m_1$，$m_2$，…，$m_k$，即$A$的特征多项式为

$$f(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_k)^{m_k}$$

则e^{tA}的表达式为

$$e^{tA}=e^{\lambda_1 t}\sum_{j=0}^{m_1-1}\frac{t^j}{j!}C_{1j}+e^{\lambda_2 t}\sum_{j=0}^{m_2-1}\frac{t^j}{j!}C_{2j}+\cdots$$

$$+e^{\lambda_k t}\sum_{j=0}^{m_k-1}\frac{t^j}{j!}C_{kj}$$

而对于本题中的A，则有

$$e^{tA}=e^{5t}C_0+e^t(C_1+tC_2)$$

两边取Laplace变换：

$$\mathscr{L}(e^{tA})=\mathscr{L}(e^{5t})C_0+\mathscr{L}(e^t)C_1+\mathscr{L}(te^t)C_2$$

$$(sI-A)^{-1}=\frac{1}{s-5}C_0+\frac{1}{s-1}C_1+\frac{1}{(s-1)^2}C_2$$

上式两端用$f(s)=|sI-A|=(s-1)^2(s-5)$去乘，便得到

$$B(s)\triangleq \mathrm{adj}(sI-A)=(s-1)^2C_0+(s-1)(s-5)C_1+(s-5)C_2$$

$$B'(s)=2(s-1)C_0+(s-5)C_1+(s-1)C_1+C_2$$

将$s=5$代入$B(s)$，$s=1$代入$B(s)$，$B'(s)$，便得到

$$B(5)=16C_0，\ B(1)=-4C_2，\ B'(1)=-4C_1+C_2$$

由此解得

$$C_0=\frac{1}{16}B(5),\qquad C_1=-\frac{1}{4}B'(1)+\frac{1}{16}B(1),$$

$$C_2=-\frac{1}{4}B(1)$$

根据定理4.7.4之例2，可计算出

$$B(5)=A^2-2A+I,$$
$$B(1)=A^2-6A+5I,$$
$$B'(1)=A-5I$$

由此计算出

$$C_0=\frac{1}{16}(A^2-2A+I)$$

$$C_1=\frac{1}{16}(A^2-10A+25I)$$

$$C_2=-\frac{1}{4}(A^2-6A+5I)$$

最后便算得

$$e^{tA}=\frac{1}{16}e^{5t}(A^2-2A+I)+\frac{1}{16}e^t(A^2-10A+25I)-\frac{1}{4}te^t(A^2-6A+5I)。$$

从下边的讨论可以知道 $A^2-6A+5I=0$，故上式可以化简成

$$e^{tA}=\frac{1}{4}e^{5t}(A-I)-\frac{1}{4}e^t(A-5I)$$

$$=\frac{1}{4}(5e^t-e^{5t})I+\frac{1}{4}(e^{5t}-e^t)A$$

我们改用谱上一致多项式的方法重做一次，作个比较。从A的特征多项式$f(\lambda)=(\lambda-1)^2(\lambda-5)$知，$\lambda=1$的代数重数为2，而由

$$I-A=\begin{pmatrix}-1 & -2 & -1\\ -1 & -2 & -1\\ -1 & -2 & -1\end{pmatrix}$$

易见 $\mathrm{rank}(I-A)=1$，因此 $\lambda=1$的几何重数为2，所以A是可对角化的。据第二章知，A的最小多项式为不同线性因子的乘积，且包含A的所有不同特征值，故

$$m_A(\lambda)=(\lambda-1)(\lambda-5)$$

这意味着$A^2-6A+5I=0$，即$A^2=6A-5I$，由此可推出

$$A^r=\alpha_0I+\alpha_1A, \qquad r\geqslant 2$$

从而可断定有

$$e^{tA}=\alpha_0(t)I+\alpha_1(t)A$$

所以可设与$e^{\lambda t}$谱上一致的多项式为

$$\alpha_0(t)+\alpha_1(t)\lambda$$

将$\lambda_1=1$，$\lambda_2=5$，代入便得到

$$\begin{cases} e^t=\alpha_0(t)+\alpha_1(t)\\ e^{5t}=\alpha_0(t)+5\alpha_1(t)\end{cases}$$

解之得

$$\alpha_0(t)=\frac{1}{4}(5e^t-e^{5t}),\qquad \alpha_1(t)=\frac{1}{4}(e^{5t}-e^t)$$

从而得到

$$e^{tA}=\frac{1}{4}(5e^t-e^{5t})I+\frac{1}{4}(e^{5t}-e^t)A$$

另外，也很容易算出A的三个独立的特征向量：

$\lambda=1$：$p_1=(-2,\ 1,\ 0)^T$，　　$p_2=(-1,0,\ 1)^T$

$\lambda=5$：$P_3=(\ 1\ ,\ \ 1\ ,\ \ 1\ )^T$

令

$$P=(p_1\ \ p_2\ \ p_3)=\begin{pmatrix}-2 & -1 & 1\\ 1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}$$

则有

$$A=P\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 5\end{pmatrix}P^{-1}\triangleq P\Lambda P^{-1}$$

从而有

$$e^{tA}=\sum_{n=0}^{\infty}\frac{t^n}{n!}A^n=\sum_{n=0}^{\infty}\frac{t^n}{n!}P\Lambda^nP^{-1}$$

$$=P\begin{pmatrix}\sum\limits_{n=0}^{\infty}\frac{t^n}{n!} & 0 & 0\\ 0 & \sum\limits_{n=0}^{\infty}\frac{t^n}{n!} & 0\\ 0 & 0 & \sum\limits_{n=0}^{\infty}\frac{(5t^n)^n}{n!}\end{pmatrix}P^{-1}$$

$$=P\begin{pmatrix}e^t & 0 & 0\\ 0 & e^t & 0\\ 0 & 0 & e^{5t}\end{pmatrix}P^{-1}$$

$$=\frac{1}{4}\begin{pmatrix}-2 & -1 & 1\\ 1 & 0 & 1\\ 0 & 1 & 1\end{pmatrix}\begin{pmatrix}e^{t} & 0 & 0\\ 0 & e^{t} & 0\\ 0 & 0 & e^{5t}\end{pmatrix}\begin{pmatrix}-1 & 2 & -1\\ -1 & -2 & 3\\ 1 & 2 & 1\end{pmatrix}$$

$$=\frac{1}{4}\begin{pmatrix}3e^{t}+e^{5t} & -2e^{t}+2e^{5t} & -e^{t}+e^{5t}\\ -e^{t}+e^{5t} & 2e^{t}+2e^{5t} & -e^{t}+e^{5t}\\ -e^{t}+e^{5t} & -2e^{t}+2e^{5t} & 3e^{t}+e^{5t}\end{pmatrix}$$

这个例子表明，计算 e^{tA} 有各种方法，但要根据A的具体情况作适当的选择。

35． 给出下述方程的显式解：

(a) $\frac{dx}{dt}=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}x$， $x(0)=\begin{bmatrix}1\\ 3\end{bmatrix}$

(b) $\frac{dx}{dt}=\begin{bmatrix}4 & 1\\ -1 & 2\end{bmatrix}x-e^{4t}\begin{bmatrix}1\\ 2\end{bmatrix}$，$x(0)=\begin{bmatrix}1\\ -1\end{bmatrix}$

36． 设 A，$B\in\mathbf{R}^{n\times n}$ 是稳定矩阵。证明：对任意给定的$C\in\mathbf{R}^{n\times n}$，矩阵方程

$$AM+MB=C$$

有唯一的解M。

提示：采取证明Lyapunov 定理所使用的方法。虽然也可把方程化成等价的线性方程组

$$(A\otimes I+I\otimes B^{T})\overline{M}=\overline{C}$$

其中$A\otimes B$为A，B的 Kronecker 积，$\overline{M}$为M的拉直。注意方程的系数矩阵为满秩的，当且仅当它的所有特征值都不等于零。而$A\otimes I+I\otimes B^{T}$的特征值为 $\lambda_i+\mu_j$，i，$j=1$，2，…，n，其中λ_i，μ_j分别为A与B的特征值。用这种方法只能证明M的存在与唯一性，却不能给出M的表达式。但在第六节末给出的解$M=-\int_0^{\infty}e^{tA}Ce^{tB}dt$并沒有证明，請给出一个证明。

37． 关于Lyapunov方程$VA+A^{H}V=W$证明下述结论：

(a) 若$W=0$，V正定，则当$A=A^{H}$时$A=0$；

(b) 若W正定，$V=V^H$，则$\det(V)\neq 0$。

38. 若V_1，V_2为两个正定的Hermite矩阵，$V_1^2-V_2^2$，是正定的，则V_1-V_2是正定的。請证明之。

提示：取$W=V_1^2-V_2^2$，构造一个适当的Lyapunov方程。

39. 证明：$M(t)=e^{B(t)}$是矩阵微分方程的初值问题

$$\frac{d}{dt}M(t)=e^{tA}M(t), \qquad M(0)=I$$

的解，其中 $B(t)=\sum_{n=0}^{\infty}\frac{t^{n+1}}{(n+1)!}A^n$。

40. 为了用迭代法求方程

$$Ax=b \qquad \det(A)\neq 0$$

的解x^*，经常把$Ax=b$改写成适于迭代的等价形式：

$$x=Bx+d$$

然后，任取初始向量x_0，构造迭代序列：

$$x_{k+1}=Bx_k+d, \qquad k=0, 1, 2, \cdots$$

试问：当B满足什么条件就能保证有

$$\lim_{n\to\infty}x_n=x^*$$

提示：$\{x_n\}$收敛到x^*的充分必要条件是B的所有特征値的模都小于1。考虑迭代序列

$$x_n=B^nx_0+(I+B+B^2+\cdots+B^{n-1})d$$

41. 证明：如果对某一与向量相容的矩阵范数$\|\cdot\|$，习题40中的B满足$\|B\|<1$，则$\{x_n\}$收敛到x^*，且有误差估计

$$\|x^*-x_n\|\leqslant\frac{\|B\|^n}{1-\|B\|}(\|d\|+\|x_0\|)$$

提示：注意$x^*=(I-B)^{-1}d$，再利用习题11-12。

第五章　特征值的估计

每个方阵 $A\in\mathbf{C}^{n\times n}$，在复数域上考虑都有 n 个特征值，它们对应复平面上的 n 个点。对于这 n 个点所在的位置，给出适当的包围或排除它的区域，便是特征值的定位或估计。给出包含特征值的范围越小，估计的便越精确。特征值的计算和估计在实用和理论上都十分重要。要精确计算特征值并非总有可能，即使在某些情况下有可能，付出的代价也会太高。而在很多应用方面往往不必精确计算特征值，而有一个粗略的估计就够了：比如在自动控制理论中，通过估计 A 的特征值是否都有负实部，便可判定系统的稳定性；与差分方法的稳定性有关的问题，就是要判定矩阵的特征值是否都落在单位圆上；和线性代数方程组迭代法求解有关的问题，需要估计矩阵的特征值是否都落在单位圆内；判定级数 $\sum_{i=0}^{\infty} c_i A^i$ 是否收敛，需要估计 A 的特征值是否都落在标量级数 $\sum_{i=0}^{\infty} c_i z^i$ 的收敛圆内（当然通过考查 A 的范数也是可以的）；在使用Neumann 展式计算 A 的广义逆时，需要估计 $A^H A$ 的特征值以确定参数 c。

对一些特殊类型的矩阵，其特征值的定位问题是很容易的：对角阵或三角阵的特征值就是其主对角元；幂等阵（$A^2=A$）的特征值非零即 1；酉阵的特征值都位于复平面的单位圆周上；Hermite 阵的特征值皆落在实轴上；反 Hermite 阵的特征值都落在虚轴上。可是对于一般复方阵，其特征值的定位问题相当困难，因为它们可以落在复平面的任何地方。反过来看，在复平面上任意指定一点，便可构造出无穷多个矩阵都以该点为其谱点。所以

用矩阵的元素或矩阵元素的简单函数，来确定特征值的位置或界，一直是数值分析学者研究的重要课题。经过近百年的努力，虽然获得了不少的结果，但远远满足不了实际的需要。

本章着重介绍一些有关的经典结论，而对于某些散布在论文中的结果，不加证明地搜录在这里，并指出参考文献。

第一节　特征值的界的估計

这一节介绍估计特征值的界的几个重要不等式。为了方便，对于给定的 $A\in\mathbf{C}^{n\times n}$，我们定义

$$B=(b_{ij})=(A+A^H)/2$$

$$C=(c_{ij})=(A-A^H)/2$$

显然，B和C分别是Hermite阵和反Hermite阵。此外，还假定A、B、C的特征值分别为$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$，$\{\mu_1, \mu_2, \cdots, \mu_n\}$，$\{i\nu_1, i\nu_2, \cdots, i\nu_n\}$，并且排列的次序满足：$|\lambda_1|\geqslant|\lambda_2|\geqslant\cdots\geqslant|\lambda_n|$，$\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n$，$\nu_1\geqslant\nu_2\geqslant\cdots\geqslant\nu_n$。

定理5.1.1（Schur不等式，1909年）　若$A=(a_{ij})\in\mathbf{C}^{n\times n}$的谱为$\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$，则有

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$$

其中等号当且仅当A是正规矩阵时成立。

证明　据第二章的Schur定理，A酉相似于上三角阵，即存在酉阵U及上三角阵T满足

$$U^HAU=T,\ U^HA^HU=T^H$$

从而有

$$U^HAA^HU=TT^H$$

$$\mathrm{tr}(AA^H)=\mathrm{tr}(U^HAA^HU)=\mathrm{tr}(TT^H)$$

注意到T的主对角元就是A的特征值，便有

$$\sum_{i=1}^{n}|\lambda_i|^2=\sum_{i=1}^{n}|t_{ii}|^2\leqslant\sum_{i=1}^{n}|t_{ii}|^2+\sum_{i\neq j}\sum|t_{ij}|^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2$$

上述不等式中的等号当且仅当 $t_{ij}=0$ $(i \neq j)$ 时成立，即当且仅当 A 酉相似于对角阵时成立。

推论5.1.2（Hirsch，1902年） A 如定理 5.1.1所设，则有

1. $|\lambda_i| \leqslant n \max_{i,j} |a_{ij}|$
2. $|\text{Re}(\lambda_i)| \leqslant n \max_{i,j} |b_{ij}|$
3. $|\text{Im}(\lambda_i)| \leqslant n \max_{i,j} |c_{ij}|$

证明 据定理5.1.1的证明可得

$$U^H BU = U^H(A + A^H)U/2 = (T + T^H)/2$$

$$U^H CU = U^H(A - A^H)U/2 = (T - T^H)/2$$

故有

$$\sum_{i=1}^{n} \left| \frac{\lambda_i - \bar{\lambda}_i}{2} \right|^2 + \sum_{j=1}^{n} \sum_{i=1}^{j-1} \frac{|t_{ij}|^2}{2} = \sum_{i=1}^{n} \sum_{j=1}^{n} |b_{ij}|^2$$

$$\sum_{i=1}^{n} \left| \frac{\lambda_i + \bar{\lambda}_i}{2} \right|^2 + \sum_{j=1}^{n} \sum_{i=1}^{j-1} \frac{|t_{ij}|^2}{2} = \sum_{i=1}^{n} \sum_{j=1}^{n} |c_{ij}|^2$$

亦即

$$\sum_{i=1}^{n} |\text{Re}(\lambda_i)|^2 \leqslant \sum_{i=1}^{n} \sum_{j=1}^{n} |b_{ij}|^2 \leqslant n^2 \max_{i,j} |b_{ij}|^2$$

$$\sum_{i=1}^{n} |\text{Im}(\lambda_i)|^2 \leqslant \sum_{i=1}^{n} \sum_{j=1}^{n} |c_{ij}|^2 \leqslant n^2 \max_{i,j} |c_{ij}|^2$$

由Schur不等式易见

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \leqslant n^2 \max_{i,j} |a_{ij}|^2$$

由以上的三个不等式显然可推知推论5.1.2成立。

推论5.1.3（Bendixson，1900年） 若 A 是实矩阵，则更有

$$|\text{Im}(\lambda_i)| \leqslant \sqrt{\frac{n(n-1)}{2}} \max_{i,j} |c_{ij}|^2$$

证明 据上述推论，已有不等式

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}|c_{ij}|^2$$

注意到当A是实矩阵时，$C=(A-A^H)/2=(A-A^T)/2$的主对角元$c_{ii}=0$ $(i=1, 2, \cdots, n)$ 故上面不等式可改写为

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{i\neq j}\sum|c_{ij}|^2 \leqslant n(n-1)\max_{i,j}|c_{ij}|^2$$

再注意到实方阵A的特征多项式是实系数多项式，其特征值若是复数，则共轭成对出现。不妨设共有 s 对复特征值，这时上面的不等式的左端为

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2 = 2\sum_{i=1}^{s}|\mathrm{Im}(\lambda_i)|^2$$

从而有

$$2\sum_{i=1}^{s}|\mathrm{Im}(\lambda_i)|^2 \leqslant n(n-1)\max_{i,j}|c_{ij}|^2$$

亦即

$$\sum_{i=1}^{s}|\mathrm{Im}(\lambda_i)|^2 \leqslant \frac{n(n-1)}{2}\max_{i,j}|c_{ij}|^2$$

这样便推出了Bendixson 的估计式：

$$|\mathrm{Im}(\lambda_i)| \leqslant \sqrt{\frac{n(n-1)}{2}}\max_{i,j}|c_{ij}|$$

定理5.1.4 A，B，C，λ_i，μ_i，ν_i如前面所设，则有

1. $\mu_n \leqslant \mathrm{Re}(\lambda_i) \leqslant \mu_1$
2. $\nu_n \leqslant \mathrm{Im}(\lambda_i) \leqslant \nu_1$

证明 设 $x \in \mathbf{C}^n$ 为A的对应于特征值 λ_i 的单位长的特征向量，即$Ax=\lambda_i x$，$\|x\|^2=x^H x=1$，则有

$$(Ax|x)=\lambda_i, \quad (A^H x|x)=\bar{\lambda}_i$$

由此可得

$$\mathrm{Re}(\lambda_i)=\left(\frac{A+A^H}{2}x\,\Big|\,x\right)=(Bx|x)$$

$$i\,\mathrm{Im}(\lambda_i)=\left(\frac{A-A^H}{2}x\,|\,x\right)=(Cx\,|\,x)$$

由于B，C皆为正规矩阵，故存在酉阵U及V使得

$$U^HBU=\mathrm{diag}(\mu_1,\ \mu_2,\ \cdots,\ \mu_n)\triangleq D_1$$

$$V^HCV=\mathrm{diag}(i\nu_1,\ i\nu_2,\cdots,\ i\nu_n)\triangleq D_2$$

从而有

$$\mathrm{Re}(\lambda_i)=(Bx\,|\,x)=(UD_1U^Hx\,|\,x)=x^HUD_1U^Hx$$
$$i\,\mathrm{Im}(\lambda_i)=(Cx\,|\,x)=(VD_2V^Hx\,|\,x)=x^HVD_2V^Hx$$

令 $y=U^Hx$，$z=V^Hx$，则据酉阵不改变向量的范数可知 $y^Hy=\sum_{i=1}^{n}|y_i|^2=x^Hx=1$，$z^Hz=\sum_{i=1}^{n}|z_i|^2=x^Hx=1$。从而有

$$\mathrm{Re}(\lambda_i)=y^HD_1y=\sum_{i=1}^{n}\mu_i|y_i|^2$$

$$i\,\mathrm{Im}(\lambda_i)=z^HD_2z=\sum_{j=1}^{n}i\nu_j|z_j|^2$$

后一等式消去i，便为

$$\mathrm{Im}(\lambda_i)=\sum_{j=1}^{n}\nu_j|z_j|^2$$

根据假定有$\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n$，$\nu_1\geqslant\nu_2\geqslant\cdots\geqslant\nu_n$，故而有

$$\mu_n=\sum_{i=1}^{n}\mu_n|y_i|^2\leqslant\sum_{i=1}^{n}\mu_i|y_i|^2\leqslant\sum_{i=1}^{n}\mu_1|y_i|^2=\mu_1$$

$$\nu_n=\sum_{i=1}^{n}\nu_n|z_i|^2\leqslant\sum_{i=1}^{n}\nu_i|z_i|^2\leqslant\sum_{i=1}^{n}\nu_i|z_i|^2=\nu_1$$

这便证明了定理5.1.4。

定理5.1.5（**Browne**，1928） 设$A\in\mathbf{C}^{n\times n}$的特征值为$\lambda_1$，$\lambda_2$，…，$\lambda_n$，奇异值为$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n$，则有

$$\sigma_n\leqslant|\lambda_i|\leqslant\sigma_1,\ i=1,\ 2,\ \cdots,\ n$$

证明 据第二章的讨论知，对 Hermite 阵A^HA来讲存在酉阵

U，使得

$$U^H A^H A U = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2) \triangleq D$$

设x为A的对应于特征值 λ_i 的单位长特征向量，即 $Ax=\lambda_i x$且 $(x|x)=\|x\|^2=1$。则有

$$(A^H Ax|x)=(Ax|Ax)=\lambda_i\bar{\lambda}_i(x|x)=|\lambda_i|^2$$

另外，设 $y=U^H x$，则$(y|y)=(x|x)=\sum_{i=1}^{n}|y_i|^2=1$，从而上面的等式又可写为

$$|\lambda_i|^2=(A^H Ax|x)=(UDU^H x|x)=(DU^H x|U^H x)$$

$$=(Dy|y)=y^H Dy=\sum_{i=1}^{n}\sigma_i^2|y_i|^2$$

又因为有 $\sigma_n^2\leqslant\sum_{i=1}^{n}\sigma_i^2|y_i|^2\leqslant\sigma_1^2$，故而有

$$\sigma_n^2\leqslant|\lambda_i|^2\leqslant\sigma_1^2$$

开方便得到

$$\sigma_n\leqslant|\lambda_i|\leqslant\sigma_1$$

以上我们是用A、$(A+A^H)/2$、$(A-A^H)/2$的元素及$A^H A$的最大最小特征值来界定A的特征值的模或它们的实虚部的模。这些不等式的实用价值是有限的。如果对矩阵 $A\in\mathbf{C}^{n\times n}$再规定以下的几个标量：

$$R_i=\sum_{j=1}^{n}|a_{ij}|,\quad R=\max_i R_i$$

$$T_j=\sum_{i=1}^{n}|a_{ij}|,\quad T=\max_j T_j$$

$$P_i=R_i-|a_{ii}|,\quad Q_j=T_j-|a_{jj}|$$

则还有以下结果：

$$|\mathrm{Im}(\lambda)|\leqslant\max_{i,j}|c_{ij}|\mathrm{ctg}\left(\frac{\pi}{2n}\right) \qquad \text{(Georg Pick)}$$

$$|\lambda|\leqslant\min(R, T) \qquad \text{(Frobenius)}$$

$$|\lambda|\leqslant\frac{R+T}{2} \qquad \text{(Browae, 1930)}$$

$$|\lambda| \leqslant \max_i \frac{R_i+T_i}{2}$$ (Parker, 1937)

$$|\lambda| \leqslant (RT)^{1/2}$$ (Farnell, 1944)

$$|\lambda| \leqslant \min(R, T)$$ (Alfred Brauer, 1946)

这些结果的证明有的在后面给出，其余的作为练习留给读者。

第二节 特征值所在区域的确定，盖尔斯果林的圆盘定理

在介绍Geršgorin的圆盘定理之前，先介绍对角占优矩阵的一个重要性质。它是Levy和Desplanques分别于1881年和1887年证明的。由于它被收入到Hadamard 1903年出版的一本著作中，所以也曾叫做Hadamard定理，其实Minkowski在1900年也证明过这个定理。

定理5.2.1 (Levy-Desplanques) 如果 $A=(a_{ij})\in \mathbf{C}^{n\times n}$ 是行对角占优的，即满足

$$|a_{ii}|>P_i \quad (i=1, 2, \cdots, n)$$

其中P_i由上节末所定义，则A是非奇异的，即

$$\det(A)\neq 0$$

证明 现采取反证法。若 $\det(A)=0$，则线性方程组 $Ax=0$ 有非零解$x=(x_1, x_2, \cdots, x_n)^T$，设

$$|x_r|=\max_i |x_i|$$

则显然 $x_r\neq 0$。那么便有

$$|a_{rr}||x_r|=\left|-\sum_{j\neq r}a_{rj}x_j\right|\leqslant\sum_{j\neq r}|a_{rj}||x_j|\leqslant|x_r|P_r$$

从而推出 $|a_{rr}|\leqslant P_r$，这与假定矛盾。

推论5.2.2 如果$A\in \mathbf{C}^{n\times n}$是列对角占优的，即满足

$$|a_{jj}|>Q_j, \quad j=1, 2, \cdots, n$$

则A是非奇异的。

后面的很多定理都是关于P_i或R_i来陈述的，显然对于Q_i或T_i也成立，請读者注意。

有了定理5.2.1，便很容易地推出著名的盖尔斯果林圆盘定理：

定理5.2.3（Geršgorin，1931） 设$A=(a_{ij})\in \mathbf{C}^{n\times n}$，则它的所有特征值都落在复平面上的$n$个圆盘

$$D_i(A)=\{z\mid |z-a_{ii}|\leqslant P_i\}\ i=1,\ 2,\ \cdots,\ n$$

的并集$\left(\bigcup\limits_{i=1}^{n} D_i(A)\right)$中。

证明 对于A的任一特征值λ_j，有$\det(\lambda_j I-A)=0$。根据定理5.2.1，矩阵$(\lambda_j I-A)$一定不是行对角占优的，即至少存在一个i，使得

$$|\lambda_j-a_{ii}|\leqslant P_i$$

成立，这表明$\lambda_j\in D_i(A)\subset\bigcup\limits_{i=1}^{n} D_i(A)$。

通常我们把$D_i(A)$ $(i=1,\ 2,\ \cdots,\ n)$叫做由A所确定的Geršgorin圆盘。这个定理只是说明A的n个特征值一定落在n个圆盘的并集内，而并不能保证每个圆盘都含有A的特征值。例如

$$A=\begin{bmatrix}-4 & -10\\ 1 & 6\end{bmatrix}$$

其特征值为$1\pm\sqrt{15}$，它们都含在圆盘$D_1=\{z\mid |z+4|\leqslant 10\}$之内，而$D_2=\{z\mid |z-6|\leqslant 1\}$则不含有$A$的特征值。值得注意的是$D_1$与$D_2$相交，$D_1\cup D_2$构成了复平面上的一个连通区域。这时我们自然会提出这样一个问题：如果A的n个圆盘的并集构成复平面上的几个互不相交的连通域，是否每个连通域都含有A的特征值？下面的定理回答了这个问题。

定理5.2.4（Geršgorin，1931） 设A的n个圆盘中有s个

圆盘构成了一个连通域G，它与其余的$n-s$个圆盘都不相交，则A的n个特征值中有且仅有s个落在G中。

证明 令$A=D+C$，其中$D=\mathrm{diag}(a_{11},\ a_{22},\ \cdots,\ a_{nn})$。作参数矩阵

$$A(t)=D+tC$$

显然$A(0)=D$，$A(1)=D+C=A$，当t从0变到1时，$A(t)$由D变到A。另外，D的n个盖尔斯果林圆盘就是A的n个盖氏圆盘的圆心。而$A(t)$（$0<t<1$）的每个盖氏圆盘都落在A的一个相应的圆盘之内且是同心的。由于$A(t)$的特征值$\lambda(t)$连续地依赖于t，所以当t由0变到1时，$A(t)$的特征值$\lambda(t)$在复平面上由各自的圆心出发画出n条连续曲线，且曲线的终点为A的n个特征值。这些曲线中的每一条要么全部落在G上，要么全部落在其余的$n-s$个圆盘的并集上。因此，由G的s个圆盘的圆心出发的$\lambda(t)$，在$t\in[0,\ 1]$时应全部落在G上。否则，由于G和其余$n-s$个圆盘都不相交，而曲线$\lambda(t)$上必有一点，不妨设为$\lambda(t_0)$，$0<t_0<1$它落在A的n个圆盘的并集之外。但$\lambda(t_0)$为$A(t_0)$的特征值，它应落在$A(t_0)$的n个圆盘的并集上，而该并集又包含在A的n个圆盘的并集中，这便与前述结论矛盾。以上的分析表明，在G上至少有S个A的特征值。但从与G不相交的其余圆盘构成的连通域来看，在G上的特征值也不得超过s个。从而便证明了在G上恰有s个特征值。

上述定理说明，孤立的盖氏圆盘中含有且只含有一个A的特征值，而两个连通的盖氏圆盘中恰含有两个特征值，但并不能保证每个圆盘都一定含有A的特征值。如果A的n个圆盘两两互不相交，则A有n个互异的特征值，从而有

推论5.2.5 如果方阵$A\in \mathbf{C}^{n\times n}$的$n$个盖尔斯果林圆盘两两互不相交，则$A$相似于对角阵。

如果A是实方阵，则还有

推论5.2.6 如果方阵$A\in \mathbf{R}^{n\times n}$的$n$个盖尔斯果林圆盘两两

互不相交，则A的特征值皆为实数。

证明 实方阵的盖氏圆盘的圆心都位于实轴上。由于这些圆盘两两互不相交，所以每个圆盘只含有一个特征值。又由于实矩阵的特征多项式是实系数多项式，故若有复根则必共轭成对出现，且在实轴的上下方对称地排列。所以若有一复特征值位于某一盖氏圆盘上，则与其共轭的复特征值也必定位于该圆盘上。这便与前面的结论矛盾，从而证明A只能有实特征值。

我们利用定理5.2.1，即对角占优矩阵是非奇异的，证明了盖尔圆盘定理。反之，用定理5.2.3也很容易证明定理5.2.1。事实上，如果A是奇异的，则它至少有一个特征值为0，据定理5.2.3，它必落在某一盖氏圆盘上，不妨设落在圆盘$D_i(A)=\{z \mid |z-a_{ii}|\leqslant P_i\}$上，则取$z$为特征值0，便有$|a_{ii}|\leqslant P_i$。这便与对角占优的条件矛盾。关于对角占优矩阵有大量的研究论文，有兴趣的读者请看有关的综述性文献，这里给出两个最明显的结论。

推论5.2.7 若$A\in \mathbf{R}^{n\times n}$满足

$$a_{ii}>P_i=\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|,\quad i=1,\ 2,\ \cdots,\ n$$

则A的所有特征值都位于复平面的右半平面。

证明 只要证明A的每个特征值λ都有正实部就行了。因为A是对角占优，故A的每个特征值都不为0．设$\lambda=a+ib\neq 0$。根据定理5.2.3，至少有一个k，使得

$$|\lambda-a_{kk}|=|a+ib-a_{kk}|\leqslant P_k<a_{kk}$$

故有

$$(a-a_{kk})^2+b^2<a_{kk}^2$$

从而有

$$0<a^2+b^2<2aa_{kk}$$

又因$a_{kk}>0$，所以$a=\mathrm{Re}(\lambda)>0$。

推论5.2.8 若$A\in \mathbf{R}^{n\times n}$满足

$$|a_{ii}|>P_i=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|,\ i=1,2,\cdots,n$$

则$\det(A)$与$\prod_{i=1}^{n}a_{ii}$有相同的符号。

证明 令$A(t)=D+tC$，$f(t)=\det(A(t))$，其中$D=\mathrm{diag}(a_{11},a_{22},\cdots,a_{nn})$，$t\in[0,1]$。则对任一$t_0\in[0,1]$，$A(t_0)$都是对角占优的，故是非奇异的。而$f(t)$是$t$的多项式函数，所以关于$t$是连续的，并且有

$$f(0)=\prod_{i=1}^{n}a_{ii},\ f(1)=\det(A)$$

若$f(0)$与$f(1)$异号，则据闭区间上连续函数的中间值定理，必存在$t_0\in(0,1)$，使得

$$0=f(t_0)=\det(A(t_0))$$

这与前面$A(t_0)$是非奇异的结论矛盾。因此证明了$\det(A)$与$\prod_{i=1}^{n}a_{ii}$同号。

第三节 圆盘定理的应用及推广

盖氏圆盘定理有很多应用与推广，下面介绍几个经典的结论。

定理5.3.1 设λ为$A\in\mathbf{C}^{n\times n}$的任一特征值，则有

$$|\lambda|\leqslant\min(R,T)$$

其中$R=\max_{i}\sum_{j=1}^{n}|a_{ij}|$，$T=\max_{j}\sum_{i=1}^{n}|a_{ij}|$

证明 设λ为A的特征值，据定理5.2.1，至少存在某个i，使得

$$|\lambda-a_{ii}|\leqslant P_i=\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|$$

即有

$$|\lambda| \leqslant P_i + |a_{ii}| = \sum_{j=1}^{n} |a_{ij}| \leqslant R$$

同理可证$|\lambda| \leqslant T$。从而有$|\lambda| \leqslant \min(R, T)$。

上述定理还可改进为

定理5.3.2 (Perron, 1933) 设 $A \in \mathbf{C}^{n \times n}$, $c_1, c_2, \cdots, c_n$为任意一组正数，且令

$$K_r = \sum_{i=1}^{n} c_i |a_{ri}| / c_r, \quad r = 1, 2, \cdots, n$$

$$K = \max_r K_r$$

则对A的任一特征值λ，都有$|\lambda| \leqslant K$。

证明 由于$c_1, c_2, \cdots, c_n$皆为正数，故对角阵

$$C = \mathrm{diag}(c_1, c_2, \cdots, c_n)$$

可逆，且A与$C^{-1}AC \triangleq B$有相同的特征值。而对于矩阵B有

$$R_r = \sum_{i=1}^{n} c_i |a_{ri}| / c_r = K_r$$

将定理5.3.1应用到B，则有$|\lambda| \leqslant K$。

在定理5.3.2中，取$c_1 = c_2 = \cdots = c_n = 1$，则得到定理5.3.1，说明定理5.3.1是定理5.3.2的特殊情况。也说明，对矩阵A作适当的相似变换，有可能缩小盖氏圆盘的半径，以提高估计的精度。例如矩阵

$$A = \begin{pmatrix} 1 & 0.02 & 0.11 \\ 0.01 & 0.9 & 0.14 \\ 0.02 & 0.01 & 0.5 \end{pmatrix}$$

的三个盖氏圆盘为

$$|z-1| \leqslant 0.13, \quad |z-0.9| \leqslant 0.15, \quad |z-0.5| \leqslant 0.03$$

如果取相似变换阵$C = \mathrm{diag}(1, 1, 10^{-1})$，则矩阵

$$B = C^{-1}AC = \begin{pmatrix} 1 & 0.02 & 0.011 \\ 0.01 & 0.9 & 0.014 \\ 0.2 & 0.1 & 0.5 \end{pmatrix}$$

的三个盖氏圆盘为

$$|z-1| \leqslant 0.031, \quad |z-0.9| \leqslant 0.024, \quad |z-0.5| \leqslant 0.3$$

这三个圆盘彼此不相交，且有两个圆盘的半径被缩小了。作为盖氏圆盘定理的一个应用，下面证明Ky Fan(樊畿)定理。为此先引入一个关于非负矩阵的基本定理。

定理5.3.3 (Wielant引理) 设 $A \in \mathbf{R}^{n \times n}$ 是一个正矩阵，即它的所有元素$a_{ij} > 0$，则

1． A的谱半径$\rho(A)$ ($\rho(A) = \max_i |\lambda_i(A)|$) 是$A$的一个正特征值，且有正的特征向量，即有 $Ax = \rho(A)x$，$x = (x_1, x_2, \cdots, x_n)^T$，$x_i > 0$，$i = 1, 2, \cdots, n$

2． A的任一其它特征值的模都小于$\rho(A)$ 。

证明 设μ是A有最大模的特征值，即 $|\mu| = \rho(A)$，且对应的特征向量为$x = (x_1, x_2, \cdots, x_n)^T \neq 0$．即$Ax = \mu x$。现定义向量$p = (|x_1|, |x_2|, \cdots, |x_n|)^T$，并证明 p是A 的关于特征值$\rho(A)$ 的正特征向量。由于$Ax = \mu x$，所以对于$1 \leqslant i \leqslant n$有

$$\mu x_i = \sum_{j=1}^{n} a_{ij} x_j$$

$$\rho(A) p_i = |\mu x_i| \leqslant \sum_{j=1}^{n} a_{ij} |x_j| = \sum_{j=1}^{n} a_{ij} p_j$$

其$p_j = |x_j|$，因此便得到 $\rho(A)p \leqslant Ap$ (这里向量或矩阵的不等式是指对应的分量都满足该不等式)，亦即 $(A - \rho(A)I)p \geqslant 0$ 。下面来证明等式成立。令 $(A - \rho(A)I)p = z$且设$z \neq 0$，则由于A是正矩阵而z是非负向量 (不为0)，故有 $Az > 0$。又由于$Ap > 0$，所以能够找到一个 $\varepsilon > 0$，使得 $Az \geqslant \varepsilon Ap$。所以根据$A(A - \rho(A)I)p = Az$可得

$$A^2 p = Az + \rho(A)Ap$$

$$\geqslant \varepsilon Ap+\rho(A)Ap$$
$$\geqslant(\varepsilon+\rho(A))Ap$$

再令$B=(\varepsilon+\rho(A))^{-1}A$，则有$BAp\geqslant Ap$，从而又可推得$BA^2p\geqslant A^2p\geqslant(\varepsilon+\rho(A))Ap$及$B^2Ap\geqslant Ap$。依此类推，对一切$n\geqslant1$有$B^nAp\geqslant Ap$。又由于矩阵$B$的谱半径为$\rho(B)=(\varepsilon+\rho(A))^{-1}\rho(A)$，因此对任意的$\varepsilon>0$，$\rho(B)<1$，故$\lim\limits_{n\to\infty}B^n=0$。从而证明了$Ap\leqslant0$。但这与$Ap>0$矛盾。所以一定有$z=0$，即等式$Ap=\rho(A)p$成立。由$Ap>0$及前面等式很容易看出$\rho(A)>0$，$p>0$。这便证明了1。

为了证明2，设λ也是一个有最大模的特征值，即$|\lambda|=\rho(A)$，并设对应的特征向量为y。仿照1的推理过程，置$q_i=|y_i|$，$q=(q_1,\cdots,q_n)^T$，则有

$$\rho(A)q_i=\left|\sum_{j=1}^{n}a_{ij}y_j\right|\leqslant\sum_{j=1}^{n}a_{ij}|y_j|=\sum_{j=1}^{n}a_{ij}q_j=(Aq)_i$$

及
$$\rho(A)q=Aq$$

从而推出关系式

$$\left|\sum_{j=1}^{n}a_{ij}y_j\right|=\sum_{j=1}^{n}a_{ij}|y_j|$$

但所有的a_{ij}都是正实数，所以上面等式表明诸y_j有相同的幅角θ，写成指数形式则有

$$y_j=|y_j|e^{i\theta},\quad j=1,2,\cdots,n$$

因此$y=e^{i\theta}q$。由于q是A的关于特征值$\rho(A)$的特征向量，而y为q的一个标量倍且是非零向量，故它也是A的关于$\rho(A)$的一个特征向量。由于y也是关于λ的特征向量，而不同特征值不能有相同的特征向量，故只有$\lambda=\rho(A)$。从而证明了A的不同于$\rho(A)$的特征值，其模一定小于$\rho(A)$。

定理5.3.4 若$A\in\mathbf{R}^{n\times n}$是一个正矩阵，则矩阵$B=[\rho(A)]^{-1}A$是幂收敛的。

证明 很显然$\rho(B)=1$。把定理5.3.4应用到正矩阵B，可

知1是B的一个特征值，而其它不等于1的特征值λ，都满足$|\lambda|<1$。下面我们先证明B是幂有界的，然后证明是幂收敛的。

设y是B关于特征值1的正特征向量，即$By=y$且$y>0$，则对一切n都有$B^n y=y$。令$y_s=\max_i y_i$，$y_t=\min_i y_i$，便有

$$y_s \geqslant y_i = \sum_{l=1}^{n} (B^n)_{il} y_l \geqslant (B^n)_{ij} y_j \geqslant (B^n)_{ij} y_t$$

对所有i，j成立。这样就得到

$$0<(B^n)_{ij} \leqslant y_s / y_t$$

即证明了B是幂有界的。这意味着B的Jordan标准形中，对应于特征值$\lambda=1$的Jordan块是一阶的，否则这个Jordan块便不是幂有界的，从而与B是幂有界的结论矛盾。故证明了B是幂收敛的。

定理5.3.5 正矩阵A的特征值$\rho(A)$是一个单根，也就是说$\rho(A)$作为$\det(\lambda I-A)$的根是一重的。

证明 首先可以看出，$\rho(A)$作为A的特征根的重数等于1作为B的特征根的重数。而据定理5.3.4的证明过程，B的Jordan标准形应为

$$J=I_k \dotplus J_{n_1}(\lambda_1) \dotplus \cdots \dotplus J_{n_l}(\lambda_l)$$

其中k为1作为B的特征根的重数，也就是对应于特征值1的（一阶）Jordan块的个数，而且$|\lambda_i|<1$，$i=1,2,\cdots,l$。我们的目的是要证$k=1$。很显然有

$$J-I=0_k \dotplus J_{n_1}(\lambda_1-1) \dotplus \cdots \dotplus J_{n_l}(\lambda_l-1)$$

而由于$J_{n_1}(\lambda_1-1) \dotplus J_{n_2}(\lambda_2-1) \dotplus \cdots \dotplus J_{n_l}(\lambda_l-1)$是满秩方阵，故容易看出

$$k=\dim \mathscr{N}(J-I)=\dim \mathscr{N}(B-I)$$

设$p>0$是B关于特征值1的特征向量，即$Bp=p$，即方程$(B-I)x=0$有非零解，故$k=\dim \mathscr{N}(B-I) \geqslant 1$。如果$k>1$，则应有

与p线性无关的$q\in \mathbf{R}^n_+$满足$Bq=q$。令$\tau=q_j/p_j=\max\limits_i(q_i/p_i)$，则有$\tau p\geqslant q$。由于$\tau p-q\neq 0$，故$B(\tau p-q)>0$，从而得到

$$\tau Bp-Bq>0$$

又由于$Bp=p$，$Bq=q$，故上式为

$$\tau p-q>0，或\tau p>q$$

这样便推出，对每个$j=1，2，\cdots，n$，都有

$$\tau p_j>q_j$$

而这与我们取$\tau=q_j/p_j=\max\limits_i(q_i/p_i)$矛盾。所以$k>1$是不可能的，从而证明了$k=1$。

把定理5.3.3与定理5.3.5合在一起，便得到经典的 Perron 定理

定理5.3.6 (Perron，1907) 若$A\in \mathbf{R}^{n\times n}$是正矩阵，则$\rho(A)$是$A$的一个单重特征值，并且对应于$\rho(A)$有正的特征向量，而$A$的其它不等于$\rho(A)$的特征值$\lambda$，都满足$|\lambda|<\rho(A)$。

1912年 Frobenius 把上述定理推广到不可约的非负矩阵，结论更为圆满（见文献[9]），所以历史上又叫做 Perron-Frobenius 定理。

定理5.3.7 (Ky Fan，1958) 设$A\in \mathbf{C}^{n\times n}$，$\lambda$为$A$的任意特征值，非负矩阵$B=(b_{ij})$满足

$$b_{ij}\geqslant |a_{ij}|，\ i，j=1，2，\cdots，n$$

则存在某个i，使得

$$|\lambda-a_{ii}|\leqslant \rho(B)-b_{ii}$$

证明 先就B是正矩阵来证。据Perron定理，有正向量x，使得$Bx=\rho(B)x$。现令

$$X=\mathrm{diag}(x_1，x_2，\cdots，x_n)$$
$$C=X^{-1}AX$$

则据盖尔斯果林圆盘定理，至少有某一i使得

$$|\lambda - a_{ii}| \leqslant \sum_{\substack{j=1 \\ j \neq i}}^{n} |c_{ij}|$$

而据C的定义可知

$$|\lambda - a_{ii}| \leqslant \left(\sum_{j=1}^{n} |a_{ij}| x_j - |a_{ii}| x_i \right) / x_i$$

$$\leqslant \left(\sum_{j=1}^{n} b_{ij} x_j - b_{ii} x_i \right) / x_i$$

据$Bx = \rho(B)x$又有$\sum_{j=1}^{n} b_{ij} x_j = \rho(B) x_i$，故而有

$$|\lambda - a_{ii}| \leqslant \rho(B) - b_{ii}$$

如果B不是正矩阵，那么对所有的i，j，令$b_{ij}^{(n)} = b_{ij} + \frac{1}{n}$，则$B_n = (b_{ij}^{(n)})$便是正矩阵。据前面的证明，对每个$n$，都存在某个$i_n$，使得

$$|\lambda - a_{i_n i_n}| \leqslant \rho(B_n) - b_{i_n i_n}^{(n)}$$

由于$1 \leqslant i_n \leqslant n$，所以当$n = 1, 2, \cdots$时，$i_n$将无限次地取1到$n$间的某个正整数$i$，把取$i$的$i_n$的下脚$n$依次记为$n_1$，$n_2$，…，$n_k$，…，于是有

$$|\lambda - a_{ii}| \leqslant \rho(B_{n_k}) - b_{ii}^{(n_k)}$$

又注意到矩阵的特征值（及其绝对值）为矩阵元素的连续函数，故而有

$$\lim_{k \to \infty} \rho(B_{n_k}) = \rho(B), \quad \lim_{k \to \infty} b_{ij}^{(n_k)} = b_{ij}$$

从而证明了

$$|\lambda - a_{ii}| \leqslant \rho(B) - b_{ii}$$

下面我们再介绍盖尔斯果林圆盘定理的一些推广。首先给出Levy-Desplanques定理的一个推广。

定理5.3.8 设$A = (a_{ij}) \in \mathbf{C}^{n \times n}$，$0 \leqslant \alpha \leqslant 1$，且令

$$P_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad Q_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}|$$

如果对每个$i=1, 2, \cdots, n$ 都有

$$|a_{ii}|>P_i^{\alpha}Q_i^{1-\alpha} \qquad (*)$$

则$\det(A)\neq 0$。

证明 现就$\alpha\in(0, 1)$ 来证，因为当$\alpha=0,1$时，该定理就是定理5.2.1及推论5.2.2。还可假定，对一切$1\leqslant i\leqslant n$都有$P_i>0$，否则由$P_i=0$可知A的第i行中的非主对角元皆为零，问题便归结到$n-1$阶方阵的情形。下面采取反证法。如果条件（*）成立而$\det(A)=0$，则方程$Ax=0$有非零解$x=(x_1, x_2, \cdots, x_n)^T$，即

$$\sum_{j=1}^{n} a_{ij}x_j=0 \quad i=1, 2, \cdots, n$$

因此有

$$|a_{ii}||x_i|=\left|\sum_{\substack{j=1\\j\neq i}}^{n} a_{ij}x_j\right|\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|, \quad i=1, 2, \cdots, n$$

但由于$|a_{ii}|>P_i^{\alpha}Q_i^{1-\alpha}$，因而有

$$P_i^{\alpha}Q_i^{1-\alpha}|x_i|\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|^{\alpha}|a_{ij}|^{1-\alpha}|x_j|, \quad i=1, 2, \cdots, n$$

并且在上面的n个不等式中至少有一个使$x_i\neq 0$的i，是严格不等的。对上式的右端施用 Hölder 不等式

$$\sum_{i=1}^{n}|y_iz_i|\leqslant\left(\sum_{i=1}^{n}|y_i|^{1/\alpha}\right)^{\alpha}\left(\sum_{i=1}^{n}|z_i|^{1/(1-\alpha)}\right)^{1-\alpha}$$

便可得到

$$\begin{aligned}P_i^{\alpha}Q_i^{1-\alpha}|x_i|&\leqslant\left(\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|\right)^{\alpha}\left(\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}\right)^{1-\alpha}\\&=P_i^{\alpha}\left(\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}\right)^{1-\alpha}\end{aligned}$$

亦即

$$Q_i|x_i|^{1/(1-\alpha)}\leqslant\sum_{\substack{j=1\\j\neq i}}^{n}|a_{ij}||x_j|^{1/(1-\alpha)}$$

对于上述的n个不等式至少有一个是严格不等的，因而对上式两

端关于 i 求和便得

$$\sum_{i=1}^{n} Q_i |x_i|^{1/(1-\alpha)} < \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| |x_j|^{1/(1-\alpha)}$$

$$= \sum_{j=1}^{n}\left(|x_j|^{1/(1-\alpha)} \sum_{\substack{i=1\\ i\neq j}} |a_{ij}|\right)$$

$$= \sum_{j=1}^{n} Q_j |x_j|^{1/(1-\alpha)}$$

这是一个矛盾的结果。所以必有$\det(A)\neq 0$。

定理5.3.9（Ostrowski 圆盘定理） 设 $A\in \mathbf{C}^{n\times n}$，$0\leqslant\alpha\leqslant 1$，$\lambda$为$A$的任一特征值。则至少有一个 i，$1\leqslant i\leqslant n$，使得

$$|\lambda - a_{ii}|\leqslant P_i^{\alpha} Q_i^{1-\alpha}$$

证明 由于λ是A的特征值，故而 $\det(\lambda I - A)=0$。于是根据定理5.3.8可知，至少有一个 i，$1\leqslant i\leqslant n$，使得

$$|\lambda - a_{ii}|\leqslant P_i^{\alpha} Q_i^{1-\alpha}$$

定理5.3.9表明，A的 n 个特征值都落在下面 n 个圆盘

$$D_i(A) = \{ z \mid |z - a_{ii}|\leqslant P_i^{\alpha} Q_i^{1-\alpha} \}，\ i=1，2，\cdots，n$$

的并集中。而盖氏圆盘定理只是取 $\alpha=1$ 的特殊情况。对角占优阵是非奇异的（定理5.2.1）这一定理还有另一种形式的推广，虽然它的条件比定理 5.3.8 强些，但比较好用，避免超越代数的运算。为了给出这个结果，先介绍一个引理：

引理5.3.10 设σ，τ为任意两个非负实数，且设$\alpha\in[0,1]$，则有

$$\tau^{\alpha}\sigma^{1-\alpha}\leqslant\alpha\tau+(1-\alpha)\sigma$$

证明 不失一般性，假定 $0\leqslant\tau\leqslant\sigma$，$\alpha\in(0，1)$，且令 $x=\tau/\sigma$，则显然$0\leqslant x\leqslant 1$。又设

$$f(x) = x^{\alpha} - \alpha x，\ x\in[0，1]$$

则由于当$x<1$时有$f'(x)=\alpha(x^{\alpha-1}-1)>0$，且有$f'(1)=0$，故$f(x)$是$[0，1]$上的单调增函数，在$x=1$取最大值，从而得到

$$x^{\alpha} - \alpha x\leqslant 1-\alpha$$

将$x=\tau/\sigma$代入上式稍作整理便得到

$$\tau^{\alpha}\sigma^{1-\alpha}\leqslant\alpha\tau+(1-\alpha)\sigma$$

定理5.3.11 设$A\in\mathbf{C}^{n\times n}$，$\alpha\in[0, 1]$。如果满足

$$|a_{ii}|>\alpha P_i+(1-\alpha)Q_i,\quad i=1, 2, \cdots, n$$

则必有$\det(A)\neq 0$。

证明 据前述引理有

$$\alpha P_i+(1-\alpha)Q_i\geqslant P_i^{\alpha}Q_i^{1-\alpha}$$

从而据定理条件知

$$|a_{ii}|>P_i^{\alpha}Q_i^{1-\alpha}$$

再应用定理5.3.8便得$\det(A)\neq 0$。

Ostrowski 圆盘定理也给出特征值界的估计。

推论5.3.12 设$A\in\mathbf{C}^{n\times n}$，$\alpha\in[0, 1]$，$\lambda$为$A$的任一特征值，则有

1. $|\lambda|\leqslant\max\limits_i\{|a_{ii}|+P_i^{\alpha}Q_i^{1-\alpha}\}$
2. $|\lambda|\leqslant\max\limits_i\{|a_{ii}|-\alpha P_i+(1-\alpha)Q_i\}$

推论5.3.13 在推论5.3.12的假设下，有

$$|\lambda|\leqslant\max_i(R_i^{\alpha}T_i^{1-\alpha})$$

其中 $R_i=\sum\limits_{j=1}^{n}|a_{ij}|$，$T_j=\sum\limits_{i=1}^{n}|a_{ij}|$。

证明 把推论5.3.12的1中的右端变形：

$$\begin{aligned}|a_{ii}|+P_i^{\alpha}Q_i^{1-\alpha}&=|a_{ii}|^{\alpha}|a_{ii}|^{1-\alpha}+P_i^{\alpha}Q_i^{1-\alpha}\\&=(|a_{ii}|^{\alpha},\ P_i^{\alpha})\begin{pmatrix}|a_{ii}|^{1-\alpha}\\Q_i^{1-\alpha}\end{pmatrix}\end{aligned}$$

再应用Hölder 不等式，便得到

$$\begin{aligned}|a_{ii}|+P_i^{\alpha}Q_i^{1-\alpha}&\leqslant(|a_{ii}|^{\alpha/\alpha}+P_i^{\alpha/\alpha})^{\alpha}(|a_{ii}|^{(1-\alpha)/(1-\alpha)}+\\&\quad+Q_i^{(1-\alpha)/(1-\alpha)})^{1-\alpha}\\&=(|a_{ii}|+P_i)^{\alpha}(|a_{ii}|+Q_i)^{1-\alpha}\\&=R_i^{\alpha}T_i^{1-\alpha}\end{aligned}$$

故有 $|\lambda| \leqslant \max_i (R_i^\alpha T_i^{1-\alpha})$。

在推论5.3.13中取 $\alpha = \frac{1}{2}$，便得到比Farnell的估计($|\lambda| \leqslant (RT)^{1/2}$) 更好的估计

$$|\lambda| \leqslant \max_i (R_i T_i)^{1/2}$$

由于算术平均值大于或等于其几何平均值：

$$(R_i + T_i)/2 \geqslant (R_i T_i)^{1/2}$$

$$(R+T)/2 \geqslant (RT)^{1/2}$$

故这一估计比Browne的估计 ($|\lambda| \leqslant (R+T)/2$) 和 Parker 的估计 ($|\lambda| \leqslant \max_i (R_i + T_i)/2$) 都要精确。

Ostrowski 还从另外一个方向推广了盖氏圆盘定理，他首先定义A的 Cassini 卵形面

$$O_{ij}(A) = \{ z \in \mathbf{C} \mid |z - a_{ii}||z - a_{jj}| \leqslant P_i P_j \}$$

其中 $1 \leqslant i$，$j \leqslant n$，$i \neq j$。

定理5.3.14 (Ostrowski) 设$A \in \mathbf{C}^{n \times n}$，$n \geqslant 2$，则$A$的每个特征值$\lambda$都位于某个$O_{ij}(A)$ 之上。

证明 由于λ为A的特征值，故$(\lambda I - A)x = 0$ 有非零解 $x = (x_1, x_2, \cdots, x_n)^T$。现选取$r$，$t$满足

$$|x_r| \geqslant |x_t| \geqslant |x_j|, \quad r \neq t, \quad j \neq r$$

如果$x_t = 0$，则x的分量除了x_r之外皆为零，因此便有

$$\lambda x_r = \sum_{j=1}^n a_{rj} x_j = a_{rr} x_r$$

因$x_r \neq 0$，故有 $\lambda = a_{rr}$，从而有关系式

$$|\lambda - a_{rr}||\lambda - a_{tt}| = 0 \leqslant P_r P_t$$

如果 $x_t \neq 0$，则对所有的 i 都有

$$(\lambda - a_{ii}) x_i = \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j$$

因此更有

$$|\lambda - a_{rr}||x_r| \leqslant \sum_{\substack{j=1 \\ j \neq r}}^{n} |a_{rj}||x_j| \leqslant |x_t| P_r$$

$$|\lambda - a_{tt}||x_t| \leqslant \sum_{\substack{j=1 \\ j \neq t}}^{n} |a_{tj}||x_j| \leqslant |x_r| P_t$$

综合上面两式，便得

$$|\lambda - a_{tt}||\lambda - a_{rr}||x_r||x_t| \leqslant P_r P_t |x_r||x_t|$$

又$|x_r| \neq 0$，$|x_t| \neq 0$，故有

$$|\lambda - a_{rr}||\lambda - a_{tt}| \leqslant P_r P_t$$

因此，不管x_t是否为零，总可找到$r \neq t$，使得

$$\lambda \in O_{rt}(A)$$

推论5.3.15 若 $A \in \mathbf{C}^{n \times n}$，$n > 2$且对所有的$i \neq j$都满足

$$|a_{ii}||a_{jj}| > P_i P_j$$

则必有$\det(A) \neq 0$。

证明 设 $|a_{ii}||a_{jj}| > P_i P_j$ 成立而 $\det(A) = 0$，则$Ax = 0$有非零解，$\lambda = 0$为A的一个特征值。由定理5.3.14知，必然存在$r \neq t$，使得

$$|a_{rr}||a_{tt}| \leqslant P_r P_t$$

这便出现了矛盾，故必有$\det(A) \neq 0$。

对于矩阵

$$A = \begin{pmatrix} 2.0 & 1.1 & 1 \\ -0.8 & 3.0 & 2 \\ 1.2 & 1.1 & 3 \end{pmatrix}$$

有$(P_1, P_2, P_3) = (2.1, 2.8, 2.3)$，故易看出

$$|a_{11}||a_{22}| = 6 > P_1 P_2$$

$$|a_{11}||a_{33}| = 6 > P_1 P_3$$

$$|a_{22}||a_{33}| = 9 > P_2 P_3$$

所以A为非奇异阵，但仅用盖氏圆盘定理是不足以判定A是非奇异的。

关于 Ostrowski 的 Cassini 卵形面定理还有更为一般的形式。为此，先证明一个引理

引理5.3.16 若 $A\in C^{n\times n}$，$\alpha\in[0,1]$，且对所有的 $i\neq j$ 满足

$$|a_{ii}||a_{jj}|>P_i^{\alpha}Q_i^{1-\alpha}P_j^{\alpha}Q_j^{1-\alpha} \qquad (*)$$

则必有 $\det(A)\neq 0$。

证明 置 $s_i=P_i^{\alpha}Q_i^{1-\alpha}/|a_{ii}|$，重新赋序可使

$$s_{m_1}\geqslant s_{m_2}\geqslant\cdots\geqslant s_{m_n}$$

而引理中的条件（$*$）等价于对一切 $i\neq j$ 都有

$$s_i s_j<1$$

特别对于 $i=m_1$，$j=m_2$ 时也有 $s_{m_1}\cdot s_{m_2}<1$，故而有 $s_{m_2}<1$。如果 $s_{m_2}=0$，则 $s_{m_3}=\cdots=s_{m_n}=0$，从而在矩阵 A 中除了第 m_1 行之外的所有非主对角元都为零，故

$$\det(A)=a_{11}a_{22}\cdots a_{nn}$$

但由于 $a_{ii}\neq 0$，$i=1,2,\cdots,n$，故 $\det(A)\neq 0$。如果 $0<s_{m_2}<1$，置 $q=\sqrt{s_{m_1}/s_{m_2}}$，并设矩阵 B 是 A 的第 m_1 行和第 m_1 列乘以 q 后所得的矩阵。为方便计，用带撇的字母表示 B 的有关的量，于是有

$$P'_{m_1}=qP_{m_1},\quad Q'_{m_1}=qQ_{m_1},\quad |b_{m_1m_1}|=q^2|a_{m_1m_1}|$$

因而有

$$s'_{m_1}=s_{m_1}/q=\sqrt{s_{m_1}s_{m_2}}<1$$

对于 $i>1$，还有（据 B 的定义可证）

$$s'_{m_i}\leqslant qs_{m_i}\leqslant qs_{m_2}\leqslant\sqrt{s_{m_1}s_{m_2}}<1$$

再根据 s'_{m_i} 的定义便有

$$s'_{m_i}=(P'_i)^{\alpha}(Q'_i)^{1-\alpha}/|b_{ii}|<1$$

从而据定理5.3.8知 B 是非奇异的。而由 B 的定义又知 $\det(B)=q^2\det(A)$，故 $\det(A)\neq 0$。

定理5.3.17 (Ostrowski) 设$A\in\mathbf{C}^{n\times n}$，$n\geqslant 2$，$\alpha\in[0, 1]$，$\lambda$为$A$的任一特征值。则$\lambda$至少落在下列$n(n-1)/2$个Cassini卵形面

$$|z-a_{ii}||z-a_{jj}|\leqslant(P_iP_j)^{\alpha}(Q_iQ_j)^{1-\alpha}$$
$$i, j=1, 2, \cdots, n, i\neq j$$

中的一个之上。

证明 如果定理结论不真，则对所有的$i\neq j$都有

$$|\lambda-a_{ii}||\lambda-a_{jj}|>(P_iP_j)^{\alpha}(Q_iQ_j)^{1-\alpha}$$

据前述引理5.3.16应有

$$\det(\lambda I-A)\neq 0$$

这与λ是A的特征值矛盾，故定理成立。

第四节 特殊类型矩陣的特征值估計

到目前为止，我们主要讨论一般复方阵的特征值的估计，其中最重要的结果应是Geršgorin圆盘定理及Ostrowski的推广，但这些定理的应用还是相当有限的。我们有理由期望对一些较为特殊的矩阵做出更进一步的结果，本节将简要地介绍这方面的一些结论。首先考虑非负矩阵。

一、非负矩阵谱半径的估计

如果一个实方阵A的所有元素都大于零，则称A为一个正矩阵；若A的每个元素都大于或等于零，且至少有一个元素大于零，则称之为非负矩阵。为了讨论的方便，我们仍规定

$$R_i=\sum_{j=1}^{n}a_{ij}, \quad r=\min_i R_i, \quad R=\max_i R_i$$

定理5.4.1 (Frobenius) 设A是非负矩阵，则有

$$r\leqslant\rho(A)\leqslant R$$

证明 我们就A是正矩阵来证明，因为对A是非负阵的情

形，可仿照 Ky Fan 定理的证明，利用连续性得到相同的结论。据 Perron 定理知，$\rho(A)$ 为 A 的一个特征值且有正特征向量 x。设

$$x=(x_1, x_2, \cdots, x_n)^T, \quad x_s=\max_i x_i, \quad x_t=\min_i x_i$$

则有

$$\sum_{j=1}^{n} a_{ij} x_j=\rho(A) x_i, \quad i=1, 2, \cdots, n$$

特别有

$$\rho(A) x_s=\sum_{j=1}^{n} a_{sj} x_j \leqslant x_s \sum_{j=1}^{n} a_{sj} \leqslant R x_s$$

由于 $x_s>0$，故有 $\rho(A) \leqslant R$。类似的推理可得 $\rho(A) \geqslant r$，从而证明了 $r \leqslant \rho(A) \leqslant R$。

Frobenius 定理是说，非负矩阵的谱半径界于 A 的最大最小行（元素）和之间，其证明是 Perron 定理的明显推论，用起来也很方便。但是有点粗糙，下面给出一个改进。

定理5.4.2 (Ledermann) 如定理5.4.1所设，若 $r<R$，则有

$$r+m(1/\sqrt{\delta}-1) \leqslant \rho(A) \leqslant R-m(1-\sqrt{\delta})$$

其中 $$m=\min_{i,j} a_{ij}, \quad \delta=\max_{R_i<R_j}(R_i/R_j)$$

证明 设 $x=(x_1, x_2, \cdots, x_n)^T$ 是 A 的对应于 $\rho(A)$ 的正特征向量。显然 x 的诸分量 x_i 不能都相等。否则，便有 $R_i=\rho(A)=R$ $(i=1, 2, \cdots, n)$，这与 $r<R$ 矛盾。仍记

$$x_s=\max_i x_i, \quad x_t=\min_i x_i$$

则有 $R_t<\rho(A)<R_s$，因此有 $R_t/R_s<1$。此外还有

$$\rho(A) x_t=\sum_{j=1}^{n} a_{tj} x_j<x_s R_t$$

$$\rho(A) x_s=\sum_{j=1}^{n} a_{sj} x_j>x_t R_s$$

从而可以得到

$$\frac{x_t}{x_s} < \frac{x_s}{x_t} \frac{R_t}{R_s}$$

或
$$\frac{x_t}{x_s} < \sqrt{\frac{R_t}{R_s}} \leqslant \sqrt{\delta}$$

于是便可推得

$$\begin{aligned}\rho(A) &= \sum_{j=1}^{n} a_{sj} x_j / x_s \\ &< a_{s1} + a_{s2} + \cdots + a_{s,t-1} + a_{st}\sqrt{\delta} + \cdots + a_{sn} \\ &= R_s - a_{st}(1 - \sqrt{\delta}) \\ &\leqslant R - m(1 - \sqrt{\delta})\end{aligned}$$

$$\begin{aligned}\rho(A) &= \sum_{j=1}^{n} a_{tj} x_j / x_t \\ &> a_{t1} + a_{t2} + \cdots + a_{ts}\frac{1}{\sqrt{\delta}} + \cdots + a_{tn} \\ &= R_t + a_{ts}\left(\frac{1}{\sqrt{\delta}} - 1\right) \\ &\geqslant r + m\left(\frac{1}{\sqrt{\delta}} - 1\right)\end{aligned}$$

综合上面的结果便证明了定理5.4.2。

后来，Ostrowski 又改进了 Ledermann 的结果。

定理5.4.3 (Ostrowski) 设A是一个正矩阵。m，r，R同定理5.4.2所设，且$r<R$，则有

$$r + m\left(\frac{1}{\sigma} - 1\right) \leqslant \rho(A) \leqslant R - m(1 - \sigma)$$

其中
$$\sigma = \sqrt{(r - m)/(R - m)}$$

证明 设$x=(x_1, x_2, \cdots, x_n)^T$为$A$的对应特征值$\rho(A)$的正特征向量。为方便起见，可设

$$1 = x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n > 0$$

因为通过对A左乘置换阵P，右乘置换阵 P^{-1} 便能实现。对任意的s，t，都有

$$\rho(A)x_s \geqslant a_{s1}+x_n\sum_{j=2}^{n}a_{sj}=a_{s1}(1-x_n)+R_s x_n$$

从而有

$$\rho(A)\geqslant\frac{x_n R_s+(1-x_n)m}{x_s} \qquad (*)$$

类似地有

$$\rho(A)x_t\leqslant\sum_{j=1}^{n-1}a_{tj}+a_{tn}x_n=R_t-(1-x_n)a_{tn}$$

从而得

$$\rho(A)\leqslant\frac{R_t-(1-x_n)m}{x_t} \qquad (**)$$

特别取$R_s=R$，$R_t=r$，则由上面的关系式可得到

$$\rho(A)\geqslant x_n R+(1-x_n)m=(R-m)x_n+m$$

$$\rho(A)\leqslant(r-m)/x_t+m$$

上两式结合一起，便有

$$x_n(R-m)+m\leqslant\rho(A)\leqslant(r-m)/x_t+m$$

$$x_n(R-m)\leqslant\rho(A)-m\leqslant(r-m)/x_t$$

$$\leqslant(r-m)/x_n$$

从而有

$$x_n\leqslant\sqrt{(r-m)/(R-m)}=\sigma$$

在式（*），（**）中，取$s=n$，$t=1$则又有

$$R_n+(x_n^{-1}-1)m\leqslant\rho(A)\leqslant R_1-(1-x_n)m$$

将$x_n\leqslant\sigma$代入上式，便有

$$R_n+(\sigma^{-1}-1)m\leqslant\rho(A)\leqslant R_1-(1-\sigma)m$$

当然，更有

$$r+(\sigma^{-1}-1)m\leqslant\rho(A)\leqslant R_1-(1-\sigma)m$$

这里之所以说Ostrowski改进了Ledermann的结果，是因为

$$\sigma=\sqrt{(r-m)/(R-m)}\leqslant\sqrt{r/R}\leqslant\max_{R_i<R_j}\sqrt{R_i/R_j}=\sqrt{\delta}$$

从而有

$$r+m(\sigma^{-1}-1)\geqslant r+m(\delta^{-1/2}-1)$$
$$R-m(1-\sigma)\leqslant R-m(1-\delta^{1/2})$$

关于非负矩阵谱半径的估计的更为精细的结果是Brauer定理。事实上，可以证明该定理是利用m，R，r来估计$\rho(A)$的最佳结果，因为确实存在达到其上下界的矩阵。

定理5.4.4（Brauer） 设A是正矩阵，r，R，m的定义如前，则有

$$r+m(h-1)\leqslant\rho(A)\leqslant R-m(1-g^{-1})$$

其中

$$g=[R-2m+\sqrt{R^2-4m(R-r)}]/2(r-m)$$
$$h=[-r+2m+\sqrt{r^2-4m(R-r)}]/2m$$

证明 不失一般性可假定A满足$R_1=R$，$R_n=r$。设B为由A的最后一行乘上$g(>1)$，最后一列乘上g^{-1}所得的矩阵。显然A与B相似，有相同的谱及谱半径。B的第i行之和为

$$\sum_{j=1}^{n}a_{ij}-a_{in}+a_{in}g^{-1}=R_i-a_{in}(1-g^{-1})$$
$$\leqslant R-m(1-g^{-1})\hat{=}K_1$$
$$i=1,\ 2,\ \cdots,\ n-1$$

而B的第n行之和为

$$\sum_{j=1}^{n}a_{nj}g-a_{nn}g+a_{nn}=gr-a_{nn}(g-1)$$
$$\leqslant gr-m(g-1)\hat{=}K_2$$

现在就定理中所定义的g来证明$K_1=K_2$。因为$K_1=K_2$等价于

$$R-m(1-g^{-1})=gr-m(g-1)$$

又等价于

$$gR-mg+m=g^2r-mg^2+mg$$

即等价于

$$g^2(r-m)-(R-2m)g-m=0$$

即要求

$$g=(R-2m+\sqrt{(R-2m)^2+4m(r-m)})/2(r-m)$$

因此，在把A相似地变换成B时取上面的g，便使B的所有行和都有上界$R-m(1-g^{-1})$，从而据 Frobenius 定理可得

$$\rho(A)\leqslant R-m(1-g^{-1})$$

为了得到$\rho(A)$的下界，首先构造一个与A相似的矩阵C，它是把矩阵A的第1行除以$h(>1)$，而把矩阵A的第1列乘上h所得到的矩阵。因此矩阵C的第1行之和为

$$\sum_{j=1}^{n}a_{1j}h^{-1}-a_{11}h^{-1}+a_{11}=Rh^{-1}+a_{11}(1-h^{-1})$$

$$\geqslant Rh^{-1}+m(1-h^{-1})\triangleq K_3$$

而矩阵C的第i行之和为

$$\sum_{j=1}^{n}a_{ij}-a_{i1}+a_{i1}h=R_i+a_{i1}(h-1)$$

$$\geqslant r+m(h-1)=K_4$$

下面证明定理中所定义的h能使$K_3=K_4$，即C的行和有一个公共的下界K_4，从而据Frobenius定理有

$$\rho(A)\geqslant r+m(h-1)$$

其实，$K_3=K_4$就是

$$Rh^{-1}+m(1-h^{-1})=r+m(h-1)$$

它等价于

$$mh^2-(r-2m)h-(R-m)=0$$

亦即

$$h=(-r+2m+\sqrt{(r-2m)^2+4m(R-m)})/2m$$

$$=(-r+2m+\sqrt{r^2+4m(R-r)})/2m$$

这正是定理中所给出的h。在推导过程中，我们要求$g>1$，

$h>1$，定理中给出的 g，h 显然是满足这个条件的。

Brauer 定理的估计比 Ostrowski 定理的估计更精确，这是显而易见的。至于能否找到这样的矩阵，使其谱半径达到定理给出的上下界，这里有文献上的例子：先假定 $R>r\geqslant nm>0$，构造矩阵

$$P=\begin{bmatrix} P_1 & P_2 \\ P_3 & m \end{bmatrix}$$

其中要求 P 的所有元都不小于 m，子矩阵 P_1 的行和都等于 $R-m$，P_2 的所有元皆为 m，而 P_3 的所有元之和为 $r-m$。则矩阵

$$B=\begin{bmatrix} P_1 & g^{-1}P_2 \\ gP_3 & m \end{bmatrix}$$

与 P 相似，且其前 $n-1$ 行中的每行元素之和都是 $R-m(1-g^{-1})=K_1$，而 B 的第 n 行之和为

$$gr-m(1-g)=K_2=K_1$$

据 Frobenius 定理有

$$\rho(B)=K_1$$

类似地可构造矩阵

$$Q=\begin{bmatrix} m & Q_2 \\ Q_3 & Q_4 \end{bmatrix}$$

其中所有的元素皆不小于 m，子矩阵 Q_4 的行和都为 $r-m$，Q_3 的所有元皆为 m，Q_2 的所有元之和为 $R-m$。则矩阵

$$C=\begin{bmatrix} m & h^{-1}Q_2 \\ hQ_3 & Q_4 \end{bmatrix}$$

与 Q 相似，且 C 的第 1 行之和为 $m+(R-m)h^{-1}=K_3$，而其余各行之和皆为 $hm+r-m=K_4=K_3$。据 Frobenius 定理有

$$\rho(C)=K_3$$

下面通过一个具体的矩阵

$$A=\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 5 \end{pmatrix}$$

来看前面几个定理所给出的估计。直接求特征值可知 $\rho(A) \doteq 7.531$，而用定理中的估计式则为：

1. Frobenius 定理

$$4 \leqslant \rho(A) \leqslant 10 \quad (\text{行})$$
$$5 \leqslant \rho(A) \leqslant 10 \quad (\text{列})$$

2. Ledermann 定理

$$4.225 \leqslant \rho(A) \leqslant 9.816 \quad (\text{行})$$
$$4.414 \leqslant \rho(A) \leqslant 9.707 \quad (\text{列})$$

3. Ostrowski 定理

$$4.732 \leqslant \rho(A) \leqslant 9.577 \quad (\text{行})$$
$$4.500 \leqslant \rho(A) \leqslant 9.667 \quad (\text{列})$$

4. Brauer 定理

$$5.162 \leqslant \rho(A) \leqslant 9.359 \quad (\text{行})$$
$$4.854 \leqslant \rho(A) \leqslant 9.472 \quad (\text{列})$$

二、稳定矩阵

在第四章我们看到，微分方程的稳定性理论中有一类重要的矩阵，它们的特征值都有负实部，并把它们叫做稳定矩阵。它所对应的一阶微分方程组 $\dot{x} = Ax$ 的零解是渐近稳定的。事实上，如果 A 的所有特征值的实部皆非正，则称其为半稳定矩阵，这种矩阵 A 与方程组 $\dot{x} = Ax$ 的零解稳定性密切相关，当它的有零实部的特征值为单重根时，则方程组的零解便是 Lyapunov 意义下稳定的。因此专门讨论这类矩阵的特征值是有特殊意义的。

定理5.4.5 若 $A \in \mathbf{R}^{n \times n}$ 的所有非主对角元 $a_{ij} \geqslant 0$，且存在一组正数 $t_1, t_2, \cdots, t_n$ 使得

$$\sum_{j=1}^{n} t_j a_{ij} \leqslant 0, \quad i = 1, 2, \cdots, n$$

则 A 是半稳定矩阵。

证明 设 λ 为 A 的特征值，$x = (x_1, x_2, \cdots, x_n)^T$ 为相应

的特征向量，且记 $y_i = x_i / t_i$，$|y_m| = \max_i |y_i|$，则有

$$\lambda x_i = \lambda t_i y_i = \sum_{j=1}^{n} a_{ij} t_j y_j, \quad i=1, 2, \cdots, n$$

特别取 $i=m$，便有

$$\lambda t_m = a_{mm} t_m + \sum_{\substack{j=1 \\ j \neq m}}^{n} a_{mj} t_j y_j / y_m$$

因此能够得到

$$|\lambda t_m - a_{mm} t_m| \leqslant \sum_{\substack{j=1 \\ j \neq m}}^{n} |t_j a_{mj}| |y_j / y_m|$$

$$\leqslant \sum_{\substack{j=1 \\ j \neq m}}^{n} t_j a_{mj}$$

再利用条件 $\sum_{j=1}^{n} t_j a_{ij} \leqslant 0$，便得到

$$|\lambda - a_{mm}| t_m \leqslant - t_m a_{mm}$$

消去 t_m，便得到 $|\lambda - a_{mm}| \leqslant - a_{mm}$。从给定的条件可知 a_{mm} 必然为负数，否则 $\sum_{j=1}^{n} t_j a_{ij} \leqslant 0$ 不可能成立。这意味着 A 的所有特征值分别位于以 $(-|a_{mm}|, 0)$ 为圆心，以 $|a_{mm}|$ 为半径的一些圆盘上，故 A 为半稳定矩阵。

从上面的证明过程可得如下推论

推论5.4.6 如果定理5.4.5中的条件 $\sum_{j=1}^{n} a_{ij} t_j \leqslant 0$ 改为 $\sum_{j=1}^{n} a_{ij} t_j < 0$ $(i=1, 2, \cdots, n)$，则 A 是稳定矩阵。

对于一般的复方阵，有 Rohrbach 定理

定理5.4.7 (Rohrbach) 若对 $A \in \mathbf{C}^{n \times n}$ 存在一组正数 t_1，t_2，…，t_n 使得

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} t_j |a_{ij}| \leqslant - t_i \mathrm{Re}(a_{ii}) \quad i=1, 2, \cdots, n$$

则 A 是半稳定矩阵。

证明 采用与定理5.4.5相同的符号，同样的推理可得

$$\lambda t_m = a_{mm} t_m + \sum_{\substack{j=1 \\ j\neq m}}^{n} a_{mj} t_j y_j / y_m$$

上式两端取实部再取模，便得

$$\begin{aligned}|\mathrm{Re}(\lambda) - \mathrm{Re}(a_{mm})| t_m &= \left|\mathrm{Re}\left(\sum_{\substack{j=1 \\ j\neq m}}^{n} a_{mj} t_j y_j / y_m\right)\right| \\ &\leqslant \left|\sum_{\substack{j=1 \\ j\neq m}}^{n} a_{mj} t_j y_j / y_m\right| \\ &\leqslant \sum_{\substack{j=1 \\ j\neq m}}^{n} |a_{mj}| t_j \leqslant - t_m \mathrm{Re}(a_{mm})\end{aligned}$$

从而得到

$$|\mathrm{Re}(\lambda) - \mathrm{Re}(a_{mm})| \leqslant - \mathrm{Re}(a_{mm})$$

这样便证明了A的特征值都有非正的实部，即A是半稳定矩阵。

与定理5.4.5的推论相仿，当条件

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} t_j |a_{ij}| \leqslant - t_i \mathrm{Re}(a_{ii}) \quad i=1, 2, \cdots, n$$

中都改为严格的不等号时，可判定A是稳定矩阵。特别是在上述条件中把所有t_j都取作1时可得

推论5.4.8 若$A \in \mathbf{C}^{n\times n}$满足

$$\mathrm{Re}(a_{ii}) \leqslant - \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \quad i=1, 2, \cdots, n$$

则A是半稳定矩阵，而当上述不等式改为严格不等式时，A为稳定性矩阵。

推论5.4.9 设$A \in \mathbf{R}^{n\times n}$满足

$$a_{ii} \leqslant - \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \quad i=1, 2, \cdots, n \qquad (*)$$

则A是稳定的当且仅当$\det(A) \neq 0$。

证明 必要性是显然的，因为稳定矩阵不能有零特征值，故A是非奇异的。现在看充分性。由于$\det(A) \neq 0$，故A不能有零

征特值。根据盖氏圆盘定理有

$$|\lambda - a_{ii}| \leqslant \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|$$

另外据条件（*）知 $\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \leqslant -a_{ii}$，故有

$$|\lambda - a_{ii}| \leqslant -a_{ii}$$

此即

$$-a_{ii} \geqslant \sqrt{(\mathrm{Re}(\lambda) - a_{ii})^2 + \mathrm{Im}^2(\lambda)} \geqslant |\mathrm{Re}(\lambda) - a_{ii}|$$

从而推出 $\mathrm{Re}(\lambda) \leqslant 0$。若 $\mathrm{Re}(\lambda) = 0$，则从上面不等式的左半部分可以判定$\mathrm{Im}(\lambda) = 0$，从而与A没有零特征值的结论矛盾。因而只能是

$$\mathrm{Re}(\lambda) < 0$$

是这就证明了A是稳定矩阵。

三、正规矩阵与Hermite矩阵

关于一般正规矩阵，这里只介绍其范数与谱半径的关系。为了比较，先给出一般方阵的结论。

定理5.4.10 方阵$A \in \mathbf{C}^{n \times n}$的谱半径不大于$A$的任何一种相容的矩阵范数。

证明 设$\|\cdot\|$是任一种相容的范数，且记$\|A\| = \alpha$。现构造矩阵$B = (\alpha + \varepsilon)^{-1} A$，其中$\varepsilon$为任一正数。很显然有

$$\|B\| = (\alpha + \varepsilon)^{-1} \|A\| = \frac{\alpha}{\alpha + \varepsilon} < 1$$

从而有

$$\|B^m - 0\| = \|B^m\| \leqslant \|B\|^m \longrightarrow 0 \quad (m \to \infty)$$

这说明$B^m \longrightarrow 0$。据第二章推论 2.2.8 知B的所有特征值的模都小于1。若A的特征值为λ，则B有特征值$\mu = (\alpha + \varepsilon)^{-1} \lambda$，从而得到

$$|\mu| = (\alpha + \varepsilon)^{-1} |\lambda| < 1$$

即有

$$|\lambda|<\alpha+\varepsilon$$

由于ε可以任意小，故得到$|\lambda|\leqslant\alpha=\|A\|$，从而证明了$A$的所有特征值的模都不大于$\|A\|$，当然有$\rho(A)\leqslant\|A\|$。

这里特别强调相容的矩阵范数是因为对某些不相容的范数，定理是不真的。例如，若取范数的定义为

$$\|A\|=\max_{i,j}|a_{ij}| \qquad (*)$$

它是不相容的，即不满足$\|AB\|\leqslant\|A\|\|B\|$。比如取

$$A=B=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$$

则有

$$AB=\begin{bmatrix}2 & 2\\ 2 & 2\end{bmatrix}$$

且$\|A\|=\|B\|=1$，$\|AB\|=2$，显然不满足

$$\|AB\|\leqslant\|A\|\|B\|$$

如果取

$$A=\begin{bmatrix}4 & 3\\ 3 & 4\end{bmatrix}$$

按（*）定义的范数有$\|A\|=4$，而$\rho(A)=7$，故定理不真。现在看正规阵的谱半径与范数间的关系。

定理5.4.11 若$A\in\mathbf{C}^{n\times n}$是正规矩阵，则有

$$\rho(A)=\|A\|_2=\max_{\|x\|_2=1}\|Ax\|_2$$

其中$\|x\|_2=\left(\sum_{i=1}^{n}|\xi_i|^2\right)^{1/2}$，$x=(\xi_1,\xi_2,\cdots,\xi_n)^T$

证明 设λ为A的一个特征值且满足$|\lambda|=\rho(A)$，x_0为对应于λ的一个特征向量且满足$\|x_0\|_2=1$。则显然有

$$\|Ax_0\|_2=\|\lambda x_0\|_2=|\lambda|\|x_0\|_2=\rho(A)$$

从而有

$$\rho(A)=\|Ax_0\|_2\leqslant\max_{\|x\|_2=1}\|Ax\|_2=\|A\|_2$$

另一方面，正规矩阵A酉相似于对角阵，即存在酉阵U，使得

$$A=U\Lambda U^H,\quad \Lambda=\text{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

再根据酉变换的保范性，故有

$$\begin{aligned}\|A\|_2&=\|U\Lambda U^H\|_2=\|\Lambda\|_2\\&=\max_{\|x\|_2=1}\|\Lambda x\|_2\\&=\max_{\|x\|_2=1}\left(\sum_{i=1}^n|\lambda_i\xi_i|^2\right)^{1/2}\\&=\max_{\|x\|_2=1}\left(\sum_{i=1}^n|\lambda_i|^2|\xi_i|^2\right)^{1/2}\\&\leqslant\max_{\|x\|_2=1}\rho(A)\left(\sum_{i=1}^n|\xi_i|^2\right)^{1/2}\\&=\rho(A)\end{aligned}$$

其中假定$x=(\xi_1,\xi_2,\cdots,\xi_n)^T$。综合以上的推导结果，便证明了$\rho(A)=\|A\|_2$。

上面的定理把$\rho(A)$的计算转换为求函数

$$f(\xi_1,\xi_2,\cdots,\xi_n)\triangleq\|Ax\|_2$$

在$\mathbf{C}^n$的单位球面上的最大值问题。引入Lagrange乘子α，还可进一步化成求函数

$$\phi(x)\triangleq x^HA^HAx+\alpha(x^Hx-1)$$

的无条件极值问题，从而成为一个微分学的问题，但解决它也不是很容易的。

Hermite矩阵作为特殊的一种正规矩阵，有更为特殊的处理方式。若$A\in\mathbf{C}^{n\times n}$是一个Hermite矩阵，其特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$皆为实数，不妨假定$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$，并令$\Lambda=\text{diag}(\lambda_1,\lambda_2,$

$\cdots$, λ_n)。则存在酉矩阵U，使得

$$A=U\Lambda U^H$$

对任给的$x\in\mathbf{C}^n$，若记$y=U^Hx$，则有

$$x^HAx=x^HU\Lambda U^Hx$$

$$=y^H\Lambda y=\sum_{j=1}^{n}\lambda_j|\eta_j|^2$$

其中假定$y=(\eta_1, \eta_2, \cdots, \eta_n)^T$。从而有

$$\lambda_ny^Hy\leqslant x^HAx\leqslant\lambda_1y^Hy$$

但由于$y^Hy=x^HUU^Hx=x^Hx$，故又有（当$x\neq0$时）

$$\lambda_n\leqslant\frac{x^HAx}{x^Hx}\leqslant\lambda_1$$

通常把x^HAx/x^Hx叫做关于A的瑞利商（Rayleigh quotient），记之为$Q_A(x)$。注意到$Q_A(x)=Q_A(x/\|x\|_2)$，因而总是把$Q_A(x)$中的x理解为单位向量加以处理。特别是把x分别取为λ_1及λ_n的特征向量x_1及x_n时，可得到$\lambda_1=Q_A(x_1)$，$\lambda_n=Q_A(x_n)$。从而有

定理5.4.12 设λ_1，λ_n分别为Hermite矩阵A的最大和最小特征值，则有

$$\lambda_1=\max_{\|x\|_2=1}Q_A(x)=\max_{x\neq0}\frac{x^HAx}{x^Hx}$$

$$\lambda_n=\min_{\|x\|_2=1}Q_A(x)=\min_{x\neq0}\frac{x^HAx}{x^Hx}$$

其实，若x_i为对应于特征值λ_i的特征向量，则也有$\lambda_i=Q_A(x_i)$。但求λ_i却不一定能化为求$Q_A(x)$的极值，所以没有什么意义。著名的Courant-Fischer定理便是上述定理的一个推广。

定理5.4.13 设Hermite矩阵A的特征值依次为：$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$，而对应的标准正交特征向量依次为$u_1$，$u_2$，$\cdots$，$u_n$，也就是说满足

$$Au_i=\lambda_iu_i，(u_i|u_j)=u_j^Hu_i=\delta_{ij}，i=1, 2, \cdots, n$$

若再记$S_0=[0]$，$S_j=[u_1, u_2, \cdots, u_j]$，$1\leqslant j\leqslant n$。则有

$$\lambda_j = \sup_{\substack{x \neq 0 \\ x \in S_{j-1}^{\perp}}} Q_A(x) = \sup\{Q_A(x) \mid x \in S_{j-1}^{\perp},\ x \neq 0\}$$

$$j = 1,\ 2,\ \cdots,\ n$$

证明 任取 $x \in \mathbf{C}^n$，由于 $\{u_1, u_2, \cdots, u_n\}$ 为 $\mathbf{C}^n$ 的一个标准正交基底，故

$$x = \sum_{l=1}^{n} \xi_l u_l,\quad \xi_l = (x \mid u_l)$$

如果 $x \in S_{j-1}^{\perp}$，则有 $\xi_1 = \xi_2 = \cdots = \xi_{j-1} = 0$，从而

$$x = \sum_{l=j}^{n} \xi_l u_l,\quad Ax = \sum_{l=j}^{n} \xi_l \lambda_l u_l$$

$$x^H A x = \sum_{i=j}^{n} \sum_{l=j}^{n} \bar{\xi}_i \xi_l \lambda_l u_i^H u_l$$

$$= \sum_{l=j}^{n} |\xi_l|^2 \lambda_l \leqslant \lambda_j \sum_{l=j}^{n} |\xi_l|^2$$

但由于 $\|x\|_2^2 = \sum_{l=1}^{n} |\xi_l|^2$，故有

$$x^H A x \leqslant \lambda_j \|x\|_2^2$$

亦即

$$Q_A(x) \leqslant \lambda_j,\quad \forall\, 0 \neq x \in S_{j-1}^{\perp}$$

特别取 $x = u_j \in S_{j-1}^{\perp}$，又有

$$Q_A(u_j) = \lambda_j$$

从而便得到

$$\lambda_j = \sup\{Q_A(x) \mid x \in S_{j-1}^{\perp},\ x \neq 0\}$$

上述定理还可以按Courant-Fischer的极小-极大形式来描述，并且能推广到求长方阵的奇异值上去（见第二章定理2.5.3）。这种考虑方式有的书上又叫做特征值的变分描述

第五节 例题与习题

1. 用Geršgorin定理证明矩阵

$$A=\begin{pmatrix} 9 & 1 & -2 & 1 \\ 0 & 8 & 1 & 1 \\ -1 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

至少有两个实特征值。

提示：A的四个 Geršgorin 圆盘为：

$$D_1=\{z\,|\,|z-9|\leqslant 4\},\quad D_2=\{z\,|\,|z-8|\leqslant 2\},$$
$$D_3=\{z\,|\,|z-4|\leqslant 1\},\quad D_4=\{z\,|\,|z-1|\leqslant 1\}$$

它们构成两个互不相接的连通域：

$$D_1\cup D_2\cup D_3,\quad D_4$$

显然，D_4 中只能包含一个特征值。由于实矩阵A的特征多项式为实系数多项式，故若有复特征值，必共轭成对出现。因此 D_4 中包含的一个特征值必为实数。而连通域$D_1\cup D_2\cup D_3$中的三个特征值至少有一个为实的，从而证明A至少有二个实特征值。

2． 用圆盘定理估计矩阵

$$A=\begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{3}{2} & i & 0 \\ 0 & -\frac{i}{2} & 5 & \frac{i}{2} \\ -1 & 0 & 0 & 5i \end{pmatrix}$$

特征值的分布范围。

3． 以二阶矩阵为例，说明圆盘定理中两个圆构成的连通部分，可以在每个圆中有一个特征值，也可以是两个特征值都落在一个圆中。

提示：取矩阵

$$A=\begin{bmatrix}2 & 1\\ 1 & 3\end{bmatrix},\ B=\begin{bmatrix}-2 & 2\\ -1 & -3\end{bmatrix}$$

$$C=\begin{bmatrix}-1 & -1\\ 1 & 1\end{bmatrix}$$

则A的特征值为$\lambda_1=\frac{1}{2}(5+\sqrt{5})$，$\lambda_2=\frac{1}{2}(5-\sqrt{5})$，其盖氏圆盘为 $D_1=\{z \mid |z-2|\leqslant 1\}$，$D_2\ \{z \mid |z-3|\leqslant 1\}$ $\lambda_1\in D_2$且$\lambda_1\notin D_1$；而$\lambda_2\in D_1$且$\lambda_2\notin D_2$。说明两个特征值分别落在两个圆盘之内。B的特征值为$\mu_1=-\frac{5}{2}+\frac{\sqrt{7}}{2}i$，$\mu_2=-\frac{5}{2}-\frac{\sqrt{7}}{2}i$；$B$的两个盖氏圆盘为$G_1=\{z \mid |z+2|\leqslant 2\}$，$G_2=\{z \mid |z+3|\leqslant 1\}$。易见$\mu_1$，$\mu_2\in G_1$，但$\mu_1$，$\mu_2\notin G_2$。这表明两个特征值落在一个圆中。$C$的两个特征值都落在两个圆盘的公共部分。

4． 证明矩阵

$$A=\begin{pmatrix}1/4 & 1/4 & 1/4 & 1/4\\ 1/5 & 2/5 & 1/5 & 1/5\\ 1/6 & 1/6 & 3/6 & 1/6\\ 1/7 & 1/7 & 1/7 & 3/7\end{pmatrix}$$

的谱半径$\rho(A)<1$，而矩阵

$$B=\begin{pmatrix}1/4 & 1/4 & 1/4 & 1/4\\ 1/5 & 2/5 & 1/5 & 1/5\\ 1/6 & 1/6 & 3/6 & 1/6\\ 1/7 & 1/7 & 1/7 & 4/7\end{pmatrix}$$

的谱半径$\rho(B)=1$。

5． 在圆盘定理中，如果一个连通部分是由两个外切圆构成的，证明每个圆上不可能有两个特征值。

6． 设U为酉阵，$A=\mathrm{diag}(a_1,a_2,\cdots,a_n)$。试证明矩阵

UA的特征值μ满足

$$m \leqslant |\mu| \leqslant M$$

其中$m=\min\limits_{1\leqslant i\leqslant n}|a_i|$，$M=\max\limits_{1\leqslant i\leqslant n}|a_i|$。

提示：设UA关于特征值μ的特征向量为x，则$UAx=\mu x$，从而可推出

$$x^H A^H A x=|\mu|^2 x^H x$$

7．证明矩阵

$$A=\begin{bmatrix} 2 & \frac{2}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & 4 & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & 2n \end{bmatrix}$$

是可对角化的，并且A的所有特征值都是实数。求出$\dim\mathscr{R}(A)$，$\dim\mathscr{N}(A)$。

提示：注意A不是对称的，但是对角占优的，且n个圆盘互相分离。据此可得到所需要的结果。

8．设A，$B\in\mathbf{R}^{n\times n}$且$B$为对称阵，作函数

$$g(A,\ B)\triangleq\frac{x^T A^T B A x}{x^T x}$$

并且令S为$\mathbf{R}^{n\times n}$中所有对称阵的集合。证明

$$\rho(A)=\min_{B\in S}\max_{x\neq 0}g(A,\ B)$$

9．设$A\in\mathbf{R}^{n\times n}$，则有

$$\lambda_i\left(\frac{A^T+A}{2}\right)\leqslant\sigma_i(A)=[\lambda_i(A^T A)]^{1/2}$$

提示：设A的奇异值分解为

$$A=UDV^T$$

其中U，V皆为正交阵。令$Q=UV^T$，$H=VDV^T$，则有

$$A=UDV^T=UV^TVDV^T=QH$$

显然Q为正交阵，H为半正定阵，这便是A的极分解。据Cauchy-Schwarz不等式有

$$y^T\left(\frac{A^T+A}{2}\right)y=y^TAy=y^TQHy\leqslant\|Hy\|\,\|Q^Ty\|$$

当$\|y\|=1$时，又有

$$\|Hy\|\,\|Q^Ty\|=\|Hy\|=(y^TH^THy)^{1/2}$$
$$=(y^TA^TAy)^{1/2}$$

这就得到不等式

$$y^T\left(\frac{A^T+A}{2}\right)y\leqslant(y^TA^TAy)^{1/2}$$

再利用定理5.4.13及定理2.5.3的结论，便可证明所需的结果。

10. 设$A\in\mathbf{C}^{n\times n}$的谱半径为$\rho(A)$，则一定存在一种相容的矩阵范数$\|\cdot\|_*$，使得对给定的$\varepsilon>0$满足

$$\rho(A)\leqslant\|A\|_*<\rho(A)+\varepsilon$$

也就是说，可以构造一种相容的矩阵范数$\|\cdot\|_*$，使得$\|A\|_*$与$\rho(A)$任意接近。这实际上也是求$\rho(A)$的一条途径。

提示：据Schur引理，存在酉阵U及上三角阵R，使得

$$U^HAU=R=(r_{ij})$$

其中$r_{ij}=0$，$i>j$；r_{ii}为A的特征值。因此

$$|r_{ii}|\leqslant\rho(A)，\quad i=1，2，\cdots，n$$

取η满足不等式

$$0<\eta\leqslant\frac{\varepsilon}{n-1}\min_{i,j}\{|r_{ij}|^{-1}|r_{ij}\neq0，j\geqslant i\}$$

不妨设$\eta<1$，令

$$D=\mathrm{diag}(1，\eta，\eta^2，\cdots，\eta^{n-1})$$

由此，可得

$$D^{-1}RD=\begin{pmatrix} r_{11} & r_{12}\eta & r_{13}\eta^2 & \cdots & r_{1n}\eta^n \\ 0 & r_{22} & r_{23}\eta & \cdots & r_{2n}\eta^{n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & r_{nn} \end{pmatrix}$$

若定义矩阵 $B\in \mathbf{C}^{n\times n}$的范数为

$$\| B \|_{*}=\| D^{-1}U^{H}BUD \|_{\infty}$$

则对任意的 B_1，B_2有

$$\begin{aligned}\| B_1B_2 \|_{*} &= \| D^{-1}U^{H}B_1B_2UD \|_{\infty} \\ &= \| D^{-1}U^{H}B_1UDD^{-1}U^{H}B_2UD \|_{\infty} \\ &\leqslant \| B_1 \|_{*}\| B_2 \|_{*}\end{aligned}$$

注意 $\| A \|_{\infty}=\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}|a_{ij}|$。这 表 明 所定义的范数 $\|\cdot\|_{*}$是相容的矩阵范数。而又有

$$\begin{aligned}\rho(A)\leqslant \| A \|_{*} &= \| D^{-1}U^{H}AUD \|_{\infty} \\ &= \| D^{-1}RD \|_{\infty} \\ &\leqslant \max_{1\leqslant i\leqslant n}|r_{ii}|+(n-1)\max_{j>i}\eta|r_{ij}| \\ &\leqslant \rho(A)+\varepsilon\end{aligned}$$

11． 设$A\in \mathbf{C}_n^{n\times n}$，$\lambda$为$A$的特征值。试证：

$$\| A^{-1} \|_2^{-1}\leqslant|\lambda|\leqslant\| A \|_2$$

12． 现假定$B\in \mathbf{C}^{n\times n}$ 与对角阵相似，且$\rho(A)<1$，则一定存在正定的Hermite阵A，使得$A-B^HAB$也正定。

13． 若A与$A-B^HAB$同为Hermite正定阵，则必有

$$\rho(B)<1$$

14． 设$A\in \mathbf{C}^{n\times n}$，$\|\cdot\|$为任一种相容的矩阵范数。则有

$$\rho(A)=\lim_{n\to\infty}\sqrt[n]{\| A^n \|}$$

提示：设A的特征值为λ_1，λ_2，…，λ_n，则A^n的特征値为λ_1^n，λ_2^n，…，λ_n^n。而又有

$$\rho(A^n)=\max_{1\leqslant i\leqslant n}|\lambda_i^n|=\max_{1\leqslant i\leqslant n}|\lambda_i|^n$$

$$=(\max_{1\leq i\leq n}|\lambda_i|)^n=[\rho(A)]^n$$

再根据矩阵的谱半径不得超过其范数的结论，又可得

$$\rho(A^n)\leqslant\|A^n\|$$

从而得到：对任意的 n，都有

$$\rho(A)=\sqrt[n]{\rho(A^n)}\leqslant\sqrt[n]{\|A^n\|}$$

当然有

$$\rho(A)\leqslant\varliminf_{n\to\infty}\sqrt[n]{\|A^n\|}\leqslant\varlimsup_{n\to\infty}\sqrt[n]{\|A^n\|}$$

只要能证明

$$\rho(A)=\varlimsup_{n\to\infty}\sqrt[n]{\|A^n\|}$$

便证明了 $\sqrt[n]{\|A^n\|}$ 的上下极限相等，都等于 $\rho(A)$。在数学分析中有众所周知的Hadamard 公式，它是说：若幂级数 $\sum_{n=0}^{\infty}c_nz^n$ 的收敛半径为 r，则对满足 $|z|<r$ 的每个 $z\in\mathbf{C}$，$\sum_{n=0}^{\infty}c_nz^n$ 都是绝对收敛的，且

$$\frac{1}{r}=\varlimsup_{n\to\infty}\sqrt[n]{|c_n|}$$

这就是所谓Hadamard 公式。下面我们利用这个公式证明

$$\rho(A)\geqslant\varlimsup_{n\to\infty}\sqrt[n]{\|A^n\|}$$

为此，对任给定的 $\varepsilon>0$，取 $\mu\in\mathbf{C}$ 满足 $|\mu|=\rho(A)+\varepsilon$，则 $A-\mu I$ 可逆。若令 $B=\dfrac{1}{\mu}A$，则 $\rho(B)<1$，故有

$$A-\mu I=-\mu(I-\mu^{-1}A)=-\mu(I-B)$$

$$(A-\mu I)^{-1}=-\frac{1}{\mu}(I-B)^{-1}$$

$$=-\frac{1}{\mu}\sum_{n=0}^{\infty}B^n=-\frac{1}{\mu}\sum_{n=0}^{\infty}\mu^{-n}A^n$$

若令 $z=\dfrac{1}{\mu}$，则

$$(A-\mu I)^{-1}=-z\sum_{n=0}^{\infty}(zA)^n$$

$$\left\|\sum_{n=0}^{\infty}z^nA^n\right\|\leqslant\sum_{n=0}^{\infty}\|A^n\||z|^n=\sum_{n=0}^{\infty}|c_n||z|^n$$

其中 $|c_n|=\|A^n\|$。当 $|z|<r$ 时，$\sum_{n=0}^{\infty}|c_n||z|^n$收敛，且

$$\frac{1}{r}=\overline{\lim_{n\to\infty}}\sqrt[n]{|c_n|}=\overline{\lim_{n\to\infty}}\sqrt[n]{\|A^n\|}$$

也就是说，由于级数对于$|z|<r$是绝对收敛的，故应有

$$|\mu|=\frac{1}{|z|}>\frac{1}{r}=\overline{\lim_{n\to\infty}}\sqrt[n]{\|A^n\|}$$

亦即

$$\rho(A)+\varepsilon>\frac{1}{r}=\overline{\lim_{n\to\infty}}\sqrt[n]{\|A^n\|}$$

由于上式对任意的ε都成立，故有

$$\rho(A)\geqslant\overline{\lim_{n\to\infty}}\sqrt[n]{\|A^n\|}$$

顺便指出，这里的证明不是很严格的。因为严格的证明需要复分析的知识。

15. 证明：每个正规矩阵$A=(a_{ij})$至少有一个特征值 λ_t，使得

$$|\lambda_t|\geqslant\max_{1\leqslant i,j\leqslant n}|a_{ij}|$$

16. 对于矩阵 $A\in\mathbf{C}^{n\times n}$，$n\geqslant3$，设其特征值为$\lambda_1$，$\lambda_2$，…，$\lambda_n$，并定义

$$s(A)=\max_{i,j}|\lambda_i-\lambda_j|$$

则有下述结论：

(a) $s(A)\leqslant\left(2\|A\|^2-\dfrac{2}{n}|\mathrm{tr}A|^2\right)^{1/2}\leqslant\sqrt{2}\;\|A\|$

(b) 若 $A \in \mathbf{R}^{n\times n}$，则

$$s(A) \leqslant \left[2\left(1-\frac{1}{n}\right)(\operatorname{tr}A)^2 - 4E_2(A)\right]^{1/2}$$

其中$E_2(A)$表示A的所有二阶主子式之和。

(c) 若$A \in \mathbf{C}^{n\times n}$ 是正规矩阵，则

$$s(A) \geqslant \sqrt{3}\max_{i\neq j}|a_{ij}|, \quad A=(a_{ij})$$

(d) 若$A \in \mathbf{C}^{n\times n}$是Hermite矩阵，则

$$s(A) \geqslant 2\max_{i\neq j}|a_{ij}|$$

$$s(A) \geqslant \max_{r\neq s}\left[(a_{rr}-a_{ss})^2 + 4|a_{rs}|^2\right]^{1/2}$$

第六章　非负矩阵及其应用

在数学经济、概率论以及系统稳定性分析等方面，常遇到一类特殊的矩阵，即所谓非负矩阵。这些矩阵的元素皆是非负的，故有某些一般矩阵所没有的公共性质。本章将介绍非负矩阵的某些性质，包括著名的Perron—Frobenius 定理，着重讨论随机矩阵以及它们在研究随机过程方面的应用。为建立起明确的概念，首先让我们看经济学中的一个例子。

例1. 四个地区 R_1，R_2，R_3，R_4 之间的某种货物的贸易关系。假定地区之间及各地区之内的货运关系用下面的分数矩阵给出

$$C=\begin{pmatrix} 0.1 & 0.3 & 0 & 0 \\ 0.7 & 0.6 & 0 & 0 \\ 0.1 & 0 & 1 & 0 \\ 0.1 & 0.1 & 0 & 1 \end{pmatrix}$$

其中c_{ij}表示地区R_j的货物每天运往地区R_i的比例。假若我们知道目前存于地区R_1，R_2，R_3，R_4的货物量分别为v_1，v_2，v_3，v_4，那么要问二个月后这些货物的分布情况是怎样的？

由于今天从R_j运往R_i的货物量是 $c_{ij}v_j$，因此明日在 R_i 的货物总量将是

$$\sum_{j=1}^{4} c_{ij}v_j$$

一般来说，$C^n v$ 给出了n 天后货物的分布情况。不难看出，把矩阵C 分块为

$$C=\begin{bmatrix} X & 0 \\ Y & I \end{bmatrix},\ X=\begin{bmatrix} 0.1 & 0.3 \\ 0.7 & 0.6 \end{bmatrix},\ Y=\begin{bmatrix} 0.1 & 0 \\ 0.1 & 0.1 \end{bmatrix}$$

便得到

$$C^n=\begin{bmatrix} X^n & 0 \\ Y\sum\limits_{j=0}^{n} X^j & I \end{bmatrix}$$

注意到X的谱半径$\rho(X)<1$，故$I-X$是可逆的，且

$$(I-X)^{-1}=\sum_{n=0}^{\infty} X^n$$

于是便有

$$\sum_{n=0}^{\infty} X^n=(I-X)^{-1}=\begin{bmatrix} 0.9 & -0.3 \\ -0.7 & 0.4 \end{bmatrix}^{-1}=\begin{bmatrix} 8/3 & 2 \\ 14/3 & 6 \end{bmatrix}$$

从而得到

$$\lim_{n\to\infty} C^n=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{4}{15} & \frac{1}{5} & 1 & 0 \\ \frac{11}{15} & \frac{4}{5} & 0 & 1 \end{pmatrix}$$

$$\lim_{n\to\infty} C^n v=\left(0,\ 0,\ \frac{4v_1+3v_2+15v_3}{15},\ \frac{11v_1+12v_2+15v_4}{15}\right)^T$$

这表明最终所有的货物都将集中到地区R_3与R_4。试问，你能从矩阵C的结构直观地了解这个结论吗？

例2. 一个随机过程的例子。为简化问题的描述，假定要考虑的物理系统总共有有限个状态，且只能在离散的时刻改变其状态。

我们用$t=0$，1，2，…表示离散的时间，而用正数1,2,…n表示系统的n个有限状态。假定系统在时刻t处于状态j，并以确定的概率m_{ij}转移到时刻$t+1$的状态i。于是可引入转移矩阵$M=(m_{ij})$，从M的定义，它显然应该满足条件

$$m_{ij} \geqslant 0, \qquad \sum_{i=1}^{n} m_{ij} = 1$$

第一个条件是说，从一种状态转移到另一种状态的概率总是非负的。第二个条件是说，系统总共有 n 个状态，某一时刻它必然处于其中某一状态，故 $\sum_{i=1}^{n} m_{ij} = 1$ 。

这两个例子中的矩阵 C 和 M，都有非负的元素，所以叫做非负矩阵。为此，我们规定：

$A \geqslant B \rightleftarrows$ 对所有的 i，j 都有 $a_{ij} \geqslant b_{ij}$

$A > B \rightleftarrows$ 对所有的 i，j 都有 $a_{ij} > b_{ij}$

A 是非负矩阵 $\rightleftarrows A \geqslant 0$

A 是正矩阵 $\rightleftarrows A > 0$

据上述规定，很容易得到结论：

$A > 0$，$0 \neq x \geqslant 0 \Longrightarrow Ax > 0$

$A > 0$，$z \geqslant w \quad \Longrightarrow Az \geqslant Aw$

且等式成立的充分必要条件是 $z = w$。

第一节　正矩阵的基本性质

关于正矩阵的一个重要性质，在第五章中通过 Wielandt 引理已经给出。为了强调它并便于后面引用，也把它列在这里。

定理6.1.1　(**Wielandt引理**)　设 $A > 0$，则

(1)　A 的谱半径 $\rho(A)$ 是 A 的具有正特征向量的一个特征值。

(2)　A 的所有其它特征值的模都小于 $\rho(A)$。

定理6.1.2　若 $n \times n$ 阶方阵 $A > 0$，则 $[\rho(A)]^{-1} A$ 是幂收敛的。

定理6.1.3　$\rho(A)$ 是正矩阵 A 的特征多项式的单根。

以上三个定理在第五章中均已给出证明。

定理6.1.4 设$n\times n$阶正矩阵A的谱半径为$\rho(A)$，定义集合

$$S(\lambda)=\{\lambda|\lambda\geqslant 0,\ 存在一个x\geqslant 0使得Ax\geqslant\lambda x\}$$

$$T(\lambda)=\{\lambda|\lambda>0,\ 存在一个x>0使得Ax\leqslant\lambda x\}$$

则
$$\rho(A)=\max_{\lambda\in S(\lambda)}\lambda=\min_{\lambda\in T(\lambda)}\lambda$$

证明 假定下面考虑的向量都是标准化的，即

$$\|x\|=\sum_{i=1}^{n}x_i=1$$

这就自然的排除了零向量。此外，还取

$$\|A\|=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}$$

因此，若$\lambda x\leqslant Ax$，便有

$$\lambda\|x\|\leqslant\|A\|\ \|x\|$$

$$0\leqslant\lambda\leqslant\|A\|$$

可见$S(\lambda)$是一个有界集合。在$A>0$时，所定义的两个集合是非空的。记

$$\lambda_0=\sup_{\lambda\in S(\lambda)}\lambda$$

并设$\{\lambda_i\}$为$S(\lambda)$中收敛到λ_0的一个序列，而$\{x^i\}$为对应的满足

$$\lambda_i x^i\leqslant Ax^i,\quad i=1,\ 2,\ \cdots$$

的向量集合。由于$\|x^i\|=1$，这表明$\{x^i\}$是$\mathbf{R}^n$中的一个有界无穷点集，据Bolzano-Weierstrass定理，$\{x^i\}$存在一个收敛的子序列，记之为$\{x^{i_k}\}$，设它收敛到x^0；从而有

$$\lambda_{i_k}x^{i_k}\leqslant Ax^{i_k}\qquad k=1,\ 2,\ \cdots$$

上式两端关于$k\to\infty$求极限，便得到

$$\lambda_0x^0\leqslant Ax^0$$

这表明$\lambda_0\in S(\lambda)$，同时也说明上确界λ_0就是最大值。下面证明λ_0为A的特征值，而x^0为相应的特征向量，也就是说不等式$\lambda_0x^0\leqslant Ax^0$为等式$Ax^0=\lambda_0x^0$。采取反证法。如若不然，不失

一般性可设

$$\sum_{j=1}^{n} a_{1j}x_j - \lambda_0 x_1 = d_1 > 0$$

$$\sum_{j=1}^{n} a_{kj}x_j - \lambda_0 x_k \geqslant 0, \quad k=2, 3, \cdots, n(*)$$

其中$x^0=(x_1, x_2, \cdots, x_n)^T$。现取

$$y = x^0 - (d_1/2\lambda_0, 0, 0, \cdots, 0)^T$$

则不难验证有

$$Ay > \lambda_0 y$$

它与$\lambda_0 = \max\limits_{\lambda \in S(\lambda)} \lambda$相矛盾，故$d_1=0$。同理可证明(*)中的不等号都是等号。从而证明了$\lambda_0$为$A$的特征值。仿照定理6.1.1的证明方法，可证得$x^0>0$。留下来的工作是要证明$\lambda_0=\rho(A)$。如果还存在一个特征值$\lambda$满足$|\lambda|\geqslant\lambda_0$，$z$为相应的特征向量，则可取$v=(|z_1||z_2|, \cdots, |z_n|)^T$，由$Az=\lambda z$便得到

$$|\lambda| v \leqslant Av$$

由此再根据λ_0的定义，应该有$|\lambda|\leqslant\lambda_0$。从而证明了$\lambda_0=|\lambda|$。从定理的证明过程可知不等式$|\lambda|v\leqslant Av$为等式。故

$$|Az| = |\lambda z| = |\lambda| v = Av$$

因此$z=Cw$，C为一复常数，$w>0$。于是$Az=\lambda z$等价于$Aw=\lambda w$，λ是正实数。这就证明了$\lambda=\lambda_0=\rho(A)$。

前面证明了$\rho(A)$的最大性，下面来证明其最小性。若对某个$y>0$，$Ay\leqslant\lambda y$，则对任意的$z\geqslant 0$，便有

$$\lambda(z|y) \geqslant (z|Ay) = (A^T z|y)$$

如果取z为A^T的相应于谱半径$\rho(A)$的特征向量，则有

$$\lambda(z|y) \geqslant (A^T z|y) = \rho(A)(z|y)$$

由于$(z|y)>0$，故$\lambda\geqslant\rho(A)$，这样便证明了

$$\rho(A) = \min_{\lambda \in T(\lambda)} \lambda$$

例1. 设$A\geqslant B>0$，则$\rho(A)\geqslant\rho(B)$。

证明 设x为A的对应于$\rho(A)$的正特征向量，则有

$$Bx \leqslant Ax = \rho(A)x$$

利用定理6.1.4中$\rho(B)$的极小性，便有

$$\rho(B) \leqslant \rho(A)$$

其实，结论对于$A \geqslant B \geqslant 0$也是成立的。

定理6.1.5 设A_n为$n \times n$阶的正矩阵，A_{n-1}表示从A_n中去掉第i(i任意）行和第i列后所得$(n-1)\times(n-1)$阶正矩阵，则有

$$\rho(A_n) > \rho(A_{n-1})$$

证明 采取反证法。记

$$\lambda_n = \rho(A_n), \quad \lambda_{n-1} = \rho(A_{n-1})$$

不失一般性，取$A_{n-1} = (a_{ij})$，$i, j = 1, 2, \cdots, n-1$，于是有

$$\sum_{j=1}^{n-1} a_{ij} y_j = \lambda_{n-1} y_i, \quad i = 1, 2, \cdots, n-1, \quad y_i > 0$$

$$\sum_{j=1}^{n} a_{ij} x_j = \lambda_n x_i, \quad i = 1, 2, \cdots, n, \quad x_i > 0$$

上面第二个等式中的前$n-1$个方程为

$$\sum_{j=1}^{n-1} a_{ij} x_j = \lambda_n x_i - a_{in} x_n$$

$$= \lambda_n (x_i - a_{in} x_n / \lambda_n) < \lambda_n x_i$$

若$\lambda_{n-1} \geqslant \lambda_n$，则上式便与$\lambda_{n-1}$的最小性

$$\lambda_{n-1} = \min_{\lambda \in T(\lambda)} \lambda$$

发生矛盾，故必有$\rho(A_n) > \rho(A_{n-1})$。

利用定理6.1.4和定理6.1.5也可证明定理6.1.3，下面给出两个证明。

证法一 采取反证法。设$\rho(A)$不是正矩阵A的单特征值。仿照对Jordan标准形的讨论可知，必存在y，使得对某个$k \geqslant 2$有

$$(A - \rho(A)I)^k y = 0, \quad (A - \rho(A)I)^{k-1} y \neq 0$$

这就意味着$(A - \rho(A)I)^{k-1} y$是A对应于特征值$\rho(A)$的一个

特征向量，故它必须是定理 6.1.4中的x^0一个标量倍，而这个倍数通过对y的适当选择，总可取为1，因此有

$$x^0=(A-\rho(A)I)^{k-1}y$$

若记
$$z=(A-\rho(A)I)^{k-2}y$$

便有
$$Az=\rho(A)z+x^0>\rho(A)z$$

仍取
$$v=(|z_1|, |z_2|, \cdots, |z_n|)^T$$

据上式便有$Av>\rho(A)v$，这与$\rho(A)$的最大性

$$\rho(A)=\max_{\lambda\in S(\lambda)}\lambda$$

发生矛盾。

证法二 设$f(\lambda)=\det(A-\lambda I)$，据行列式的求导法则，有

$$f'(\lambda)=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial|A-\lambda I|}{\partial(A-\lambda I)_{ij}}\frac{d(A-\lambda I)_{ij}}{d\lambda}$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ij}\frac{d(A-\lambda I)_{ij}}{d\lambda}$$

$$=-\sum_{i=1}^{n}X_{ii}$$

其中X_{ij}为$(A-\lambda I)_{ij}$的代数余子式。若记A_i为A阵去掉第i行和第i列所得到的子矩阵，则有

$$f'(\lambda)=-(|A_1-\lambda I|+|A_2-\lambda I|+\cdots+|A_n-\lambda I|)$$

据定理6.1.5，对所有的i，有$\rho(A)>\rho(A_i)$，又诸$|A_i-\lambda I|$为具有相同首项系数的λ-多项式，故它们在$\lambda=\rho(A)$处有相同的符号（为什么？），因此$f'(\rho(A))\neq 0$，这意味着$\rho(A)$为A的单特征值。

定理6.1.4 用最大最小性定义了正矩阵A的谱半径$\rho(A)$，其实$\rho(A)$还可等价地表述为：

定理6.1.6 设A为$n\times n$阶正矩阵，则

$$\rho(A)=\max_{x}\min_{i}\sum_{j=1}^{n}a_{ij}x_j/x_i$$

$$=\min_{x} \max_{i} \sum_{j=1}^{n} a_{ij} x_j / x_i$$

其中x在所有非负不为零的向量中变化。

这个定理的证明留给读者完成。对于正矩阵的谱半径我们有了两种表示方法，其对应的正特征向量又如何得到呢？下面的定理给出了一个解答。

定理6.1.7 设A为正矩阵，c为任一非负向量，则极限

$$v=\lim_{k\to\infty} A^k c/\rho(A)^k$$

存在，且是A对应于$\rho(A)$的一个正特征向量。除了因c的选取所决定的一个标量倍数外，它是唯一的。

证明 令$B=[\rho(A)]^{-1}A$.根据定理6.1.2，B是幂收敛的，故极限

$$v=\lim_{k\to\infty} B^k c$$

存在。v是A对应于特征值$\rho(A)$的特征向量也是显然的，因为有

$$Bv=\lim_{k\to\infty} BA^k c/\rho(A)^k=\lim_{k\to\infty} A^{k+1}c/\rho(A)^{k+1}=v$$

从而有

$$Av=\rho(A)v$$

由于我们并不计较v的一个标量倍数，不失一般性，故可假定所考虑的c满足

$$c\geqslant 0, \qquad \sum_{i=1}^{n} c_i=1$$

设y，z分别为取初值c，d时所得到的极限。选取标量t使得$y-tz$至少有一个分量为零，其余分量皆大于或等于零。在定理6.1.2的证明中曾经指出，如若$y-tz\neq 0$，必有$B(y-tz)>0$。但另一方面又有

$$B(y-tz)=By-tBz=y-tz$$

除非$y-tz=0$，否则便发生矛盾。又由于

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i = 1$$

故有$t=1$，从而证明$y=z$。

关于正矩阵还有如下的重要结论：

定理6.1.8 设A为正矩阵，x是A的对应于特征值$\rho(A)$的特征向量，y为A^T的对应于特征值$\rho(A)$的任一正特征向量。则

$$\lim_{k\to\infty}(\rho(A)^{-1}A)^k=(y^Tx)^{-1}xy^T$$

证明 仍记$B=\rho(A)^{-1}A$，则$\lim\limits_{k\to\infty}B^k$存在，记之为$L$。又由于

$$L=\lim_{k\to\infty}B^k=B\lim_{k\to\infty}B^{k-1}=BL$$

故有 $$L_i=BL_i,\quad i=1,\ 2,\ \cdots,\ n$$

其中L_i为方阵L的第i列。这表明L的各列都是B的对应于特征值1的特征向量（若L_i不是零向量的话），由于正矩阵B的特征值$\rho(B)=1$是单特征值，且B是幂收敛的，故B对应于特征值$\rho(B)=1$的特征子空间是一维的。由于

$$Bx=\rho(A)^{-1}Ax=x$$

故x也是B的对应于特征值$\rho(B)=1$的特征向量，这就意味着每个L_i皆是x的一个标量倍，不妨记之为$L_i=r_ix$，并记$r=(r_1,\ r_2,\ \cdots,\ r_n)^T$，则

$$\begin{aligned}L&=[L_1L_2\cdots L_n]\\&=[r_1xr_2x\cdots r_nx]=xr^T\end{aligned}$$

又由于$(B^T)^ky=y$，故有$y^TL=y^T$。从而有

$$y^T=y^TL=y^Txr^T$$

显而易见$y^Tx\neq0$，$r^T=(y^Tx)^{-1}y^T$。最后便得到

$$\lim_{k\to\infty}(\rho(A)^{-1}A)^k=L=(y^Tx)^{-1}xy^T$$

例2. 现在用上述结论来研究本章开头所给例1中的问题。假定有α个单位的商品按矩阵

$$C=\begin{pmatrix}0.1 & 0.2 & 0.1 & 0.1\\ 0.7 & 0.6 & 0.1 & 0.1\\ 0.1 & 0.1 & 0.7 & 0.1\\ 0.1 & 0.1 & 0.1 & 0.7\end{pmatrix}$$

在各个地区之间运输，其中c_{ij}表示每天从R_j运往R_i的现存貨物的比例。还假定这α个单位的商品在R_1，R_2，R_3，R_4之间分配，目前分别各拥有v_1，v_2，v_3，v_4个单位。试给出半月或一个月之后，这些商品分配的情况。

我们已经知道，$C^k v$给出了k天后的分配情况。我们用定理6.1.8的结论，考查一下当k很大时，$C^k v$将是怎样一个结果。若取

$$y=(1,\ 1,\ 1,\ 1)^T$$

则易见$C^T y=y$，说明C有特征值1，故知$\rho(C)\geqslant 1$，但从特征值的估计中知道

$$\rho(C)\leqslant\max_j\left(\sum_{i=1}^{4}C_{ij}\right)=1$$

因此$\rho(C)=1$。读者也可验证，取$x=(6,\ 16,\ 11,\ 11)^T$时，有$Cx=x$。从而有

$$\begin{aligned}\lim_{k\to\infty}C^k v&=(y^T x)^{-1}xy^T v\\&=\frac{\alpha}{44}(6,\ 16,\ 11,\ 11)^T\end{aligned}$$

值得注意的是，上面的极限仅取决于矩阵C及α，而与v无关。这一结论表明：不管开始时商品是如何分配的，最终地区R_i将拥有$\frac{\alpha x_i}{44}$个商品。若开始时地区R_3沒有分配到商品，或者拥有全部α个单位商品，则最终它总是有$\alpha/4$的商品。与例1比较可能会感到不好理解。在例1的情形下，若地区R_3原来沒有貨物，最终它也会有$(4v_1+3v_2)/15$个单位的貨物；如果它一开始拥有全部貨物，便永远拥有这批貨物，这与矩阵C的结构有关。

第二节 素矩阵(primitive matrix)

上一节我们简单地介绍了正矩阵的某些性质。现在讨论更为一般的非负矩阵。

定义6.2.1 对于非负矩阵 $A(A\geqslant 0)$，如果存在自然数 m，使得A^m为正矩阵，则称A是一个素矩阵。

例如
$$A=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$
便是一个素矩阵，因为$A^4>0$。从素矩阵的定义可知，所有正矩阵的集合是素矩阵集合的一个子集。下述推论是显然的。

推论6.2.2 设$A\geqslant 0$，$B\geqslant 0$，且A为素矩阵，则有

(1) A^T也是素矩阵；

(2) 对任一自然数l，A^l是素矩阵；

(3) $A+B$也是素矩阵。

证明 (1) 与(2) 是显而易见的。由于A是素矩阵，故存在$m\in N$，使得$A^m>0$。因为
$$(A+B)^m=A^m+C$$
其中$C\geqslant 0$，故 $(A+B)^m>0$，从而证明了(3) 。

既然正矩阵是素矩阵的子集，那么就会问：素矩阵是否也具有正矩阵的某些性质？下面给出回答。

定理6.2.3 设A为素矩阵。则 $\rho(A)$为A具有正特征向量的单特征值，且A的其它特征值λ都满足$|\lambda|<\rho(A)$。

证明 据定义，存在$m\in N$使得$A^m>0$，根据定理 6.1.1，存在$x>0$满足
$$A^m x=\rho(A^m)x$$
现令
$$\hat{x}=\sum_{i=0}^{m-1}\rho(A)^{-i}A^i x$$

则有
$$A\hat{x}=\sum_{i=0}^{m-1}\rho(A)^{-i}A^{i+1}x$$
$$=\sum_{i=0}^{m-2}\rho(A)^{-i}A^{i+1}x+\rho(A)^{1-m}A^m x$$
$$=\sum_{i=1}^{m-1}\rho(A)^{1-i}A^i x+\rho(A)^{1-m}\rho(A^m)x$$
由于对任意的$k\in N$，都有
$$\lambda(A^k)=(\lambda(A))^k$$
故上式又可写成
$$A\hat{x}=\rho(A)\left[\sum_{i=1}^{m-1}\rho(A)^{-i}A^i x+x\right]$$
$$=\rho(A)\sum_{i=0}^{m-1}\rho(A)^{-i}A^i x=\rho(A)\hat{x}$$
又因$\hat{x}$是正向量与非负向量之和，故$\hat{x}>0$。这便证明了$\rho(A)$是A的具有正特征向量的特征值。

注意到x和$\hat{x}$都是A^m的对应于单特征值$\rho(A)^m$的特征向量，所以x与$\hat{x}$是线性相关的。因此x也是A的对应于$\rho(A)$的正特征向量。

下面证明$\rho(A)$是A的单特征值。很显然，$\rho(A)$作为A的特征值的重数（即代数重数）l，不得超过$\rho(A)^m$作为A^m的特征值的重数。由于$\rho(A)^m$是A^m的单特征值，故$l=1$。

最后证明A的其余特征值λ都满足
$$|\lambda|<\rho(A)$$
采取反证法。若A有特征值λ满足$|\lambda|=\rho(A)$，则对某个非零的y有$Ay=\lambda y$，从而有
$$A^m y=\lambda^m y$$
$$|\lambda^m|=|\lambda|^m=\rho(A)^m=\rho(A^m)$$
由于A^m的其它所有特征值$\mu\neq\rho(A^m)$且$\rho(A^m)>|\mu|$，故必有

$\lambda^m=\rho(A^m)$，否则会导出矛盾。于是

$$A^m y=\rho(A^m)y$$

但$\rho(A^m)$为A^m的单根，故y与x线性相关。所以y也是A的对应于$\rho(A)$的特征向量，这只有当$\lambda=\rho(A)$时才可能。

定理6.2.4　设A是素矩阵，x为A的对应于$\rho(A)$的正特征向量，y是A^T的对应于$\rho(A)$的任一正特征向量，则有

$$\lim_{k\to\infty}(\rho(A)^{-1}A)^k=(y^Tx)^{-1}xy^T$$

证明　略。

读者可能会问：给定非负矩阵A，除了定义外，还有什么方法检验A是否为素矩阵？使$A^m>0$的自然数m有什么特性？为此，我们引入所谓素性指数(*index of primitivity*)的概念。

定义6.2.5　素矩阵A的素性指数$\gamma(A)$是指使$A^m>0$的最小正整数m。

关于素性指数$\gamma(A)$有下面的结论：

定理6.2.6　若非负矩阵A为素矩阵，并且对某个正整数h，和矩阵

$$A+A^2+\cdots+A^h$$

至少有$d(>0)$个正对角元，则有

$$\gamma(A)\leqslant n-d+(n-1)h$$

其中n为A的阶数。

这个定理和下面定理6.2.8的证明，需要稍为深入的非负矩阵的知识，这里不再给出。有兴趣的读者可阅读A.Berman与R.J.Plemmons合著的《*Nonnegative Matrices in the Mathematical Sciences*》。

利用定理6.2.6可以估计A的素性指数$\gamma(A)$的大小。例如，对于前面曾提到过的非负矩阵

$$A=\begin{pmatrix}0&1&1\\1&0&0\\1&1&1\end{pmatrix}$$

如果和矩阵就取A，即取$h=1$，则对角元有一个大于零，故$d=1$。而$n=3$，因此求得

$$\gamma(A)\leqslant 3-1+1\times(3-1)=4$$

实际上$A^4>0$，恰好取了不等式的上界。

推论6.2.7 设$A\geqslant 0$为$n\times n$阶的素矩阵，并且a_{ij}与a_{ji}同时为正或同时为零，则有

$$\gamma(A)\leqslant 2(n-1)$$

证明 首先可推出A^2的对角元全为正的，从而有$d=n$，$h=2$。据定理6.2.6有

$$\gamma(A)\leqslant n-n+2(n-1)=2(n-1)$$

定理6.2.8 设$A\geqslant 0$为$n\times n$阶的素矩阵，则

$$\gamma(A)\leqslant (n-1)^2+1$$

例 $n\times n$阶矩阵

$$A=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

是一个素矩阵，其素性指数$\gamma(A)=(n-1)^2+1$。这表明一般情况下定理6.2.8中的估计不能再改进了。当然，特殊情况例外。

正矩阵集合是素矩阵集合的子集，素矩阵虽然不具有正矩阵的全部性质，但却有重要的定理 6.2.3 所描述的性质。而素矩阵集合又是非负矩阵集合的子集。试问：非负矩阵有定理 6.2.3所描述的性质吗？請看非负矩阵

$$A=\begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

显然$\rho(A)=2$，A对应于$\rho(A)$的特征向量为

$$x=(\sigma,\ \sigma,\ \tau,\ 0)^T$$

这说明A没有相应于$\rho(A)=2$的正特征向量，且$\rho(A)$也不是A的单特征值。并且A的异于$\rho(A)$的特征值$\lambda=-2$可以满足$|\lambda|=\rho(A)$。这些结论几乎把定理6.2.3整个推倒。不过还保留了两点：$\rho(A)$仍然是非负矩阵A的特征值；A仍有关于$\rho(A)$的非负特征向量。这两点结论不仅对上面给定的A是成立的，而且对一般的非负矩阵也是成立的。

定理6.2.9 设$A\geqslant 0$，则$\rho(A)$是A的具有非负特征向量的特征值。

证明 设$B_k=A+\frac{1}{k}E$，E的所有元素都等于1。对任意的$k\geqslant 1$，都有

$$0\leqslant A<B_{k+1}<B_k$$

因此不难证明有

$$\rho(A)\leqslant\rho(B_{k+1})\leqslant\rho(B_k)$$

若取$\lambda=\inf\limits_{k\geqslant 1}\rho(B_k)$，则有

$$\rho(A)\leqslant\lambda=\lim_{k\to\infty}\rho(B_k)$$

由于B_k是一个正矩阵，故存在对应于特征值$\rho(B_k)$的正特征向量y^k，即

$$B_k y^k=\rho(B_k)y^k$$

如同前面所做的一样，不妨设y^k满足

$$\|y^k\|=\sum_{i=1}^n y_i^k=1$$

于是可从有界无穷点列$\{y^k\}$中取一收敛的子序列$\{y^{k_i}\}$，且设它收敛于$0\neq y^0\geqslant 0$。因此有

$$\lim_{i\to\infty}y^{k_i}=y^0$$

$$\lim_{i\to\infty}B_{k_i}y^{k_i}=\lim_{i\to\infty}B_{k_i}y^0=Ay^0$$

$$\lambda y^0=\lim_{i\to\infty}\rho(B_{k_i})\lim_{i\to\infty}y^{k_i}=\lim_{i\to\infty}\rho(B_{k_i})y^{k_i}=Ay^0$$

这表明 λ 为 A 的具有非负特征向量 y^0 的特征值，又因为 $|\lambda| \leqslant \rho(A)$，故必有 $\lambda = \rho(A)$。

下面看一个素矩阵的应用例子。考虑五个球队的循环赛，每个队与其它队都进行一場比赛，比赛成绩用一个得胜率矩阵

$$F = \begin{pmatrix} 0 & 0.6 & 0.7 & 0.5 & 0 \\ 0.4 & 0 & 0 & 0.1 & 0.3 \\ 0.3 & 1 & 0 & 0.7 & 0.2 \\ 0.5 & 0.9 & 0.3 & 0 & 0.2 \\ 1 & 0.7 & 0.8 & 0.8 & 0 \end{pmatrix}$$

来表示，其中 f_{ij} 表示过去 i 队击败 j 队的频率。例如 $f_{34} = 0.7$，意味着 3 队与 4 队的比赛中，3 队获胜率为70%。我们一方面要对各队之间的输赢制定一种赏罚规则，譬如要求 j 队付给击败它的 i 队 x_j 元；另一方面又希望这种赏罚是公正的，也就是每一队的期望净胜量不要是负的。这种方案可行吗？让我们作一个分析。j 队预计输掉的錢为 $\sum_{i=1}^{5} f_{ij} x_j$，而 $\sum_{i=1}^{5} f_{ji} x_i$ 则为其预计赢得的錢。要满足公正性，关系式

$$\sum_{i=1}^{5} f_{ji} x_i = \sum_{i=1}^{5} f_{ij} x_j$$

应对所有的 j 及某个正向量 x 成立。若取

$$a_{ji} = f_{ji}/\gamma_j, \quad \gamma_j = \sum_{i=1}^{5} f_{ij}$$

对某一 $x > 0$ 应有

$$Ax = x$$

不难验证所定义的 $A = (a_{ij})$ 是一个素矩阵，且 $\rho(A) = 1$，因此据前面的定理6.2.3，这个问题是有解的，并且除了相差一个标量倍数外，解还是唯一的。

第三节　不可约矩阵

本节介绍不可约的非负矩阵，它包含了素矩阵，并且几乎有定理 6.2.3 所描述的性质。这种矩阵有一个很实用的性质，就是易于预测其高次幂的性态。

定义6.3.1　对于矩阵$A\in C^{n\times n}$，如果存在置换阵P，使得

$$PAP^T=\begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

其中B，D 为方阵；或者当 $n=1$时 $A=0$，则称 A是可约的 (*reducible*)。否则，便称A是不可约的 (*irreducible*)。不可约矩阵有时也叫做不可分解阵 (*indecomposable matrix*)。

例如矩阵

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 4 & 5 & 6 \end{pmatrix}$$

是可约的，因为

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}=\begin{pmatrix} 0 & 0 & 0 \\ 2 & 1 & 3 \\ 5 & 4 & 6 \end{pmatrix}$$

相当于定义 6.3.1 中的 $B=[0]$，D $\begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$。易见矩阵 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$是不可约的，因为任何 2×2 阶的置换阵 P 都不能使 $P\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}P^T$成为$\begin{bmatrix} b & o \\ c & d \end{bmatrix}$的形式。

从定义6.3.1　直接可得到

推论6.3.2　(1)　当且仅当A^T是不可约时 $A\in C^{n\times n}$ 是不可约的。

(2) 若$A\geqslant 0$是不可约的，$B\geqslant 0$，则$A+B$也是不可约的。

证明 (1) 是显然的。对(2) 采取反证法。若$A+B$是可约的，则存在置换阵P使得

$$P(A+B)P^T=PAP^T+PBP^T=\begin{bmatrix} E & O \\ C & D \end{bmatrix}$$

由于A，B都是非负矩阵，故PAP^T，PBP^T都必须具有上式右端的形式，这与A是不可约的相矛盾。

直接按定义判别$A\in \mathbf{C}^{n\times n}$是否可约，需要用$n!$个置换阵$P$进行验算。当$n$很大时，这是一项计算量极大的工作。不过，对于我们有兴趣的非负矩阵来说，存在着有效的方法。若记

$$a_{ij}^{(q)}=(A^q)_{ij}$$

则有如下判别定理。

定理6.3.3 $n\times n$阶非负矩阵A是不可约的，其充分必要条件是：对所有的(i,j)，存在自然数q，使得$a_{ij}^{(q)}>0$。

证明 必要性。采取反证法。若A是不可约的，且对所有(i,j)不存在一个自然数q使得$a_{ij}^{(q)}>0$。也就是说，至少存在一对(i,j)，使得对一切自然数q都有$a_{ij}^{(q)}=0$。我们取集合

$$S=\bigcup_{q\geq 1}\{l \mid a_{lj}^{(q)}>0\},\quad T=\bigcup_{q\geq 1}\{l \mid a_{il}^{(q)}>0\}$$

由于有

$$a_{ij}^{(q+p)}\geqslant a_{il}^{(q)}a_{lj}^{(p)}$$

故有$S\cap T=\phi$。若$S=\phi$，则A有一列为零向量；若$T=\phi$，则A有一行为零向量。这两种情形都表明A是可约的，与假设A是不可约的矛盾。现考察S与T皆不空的情况。因为

$$a_{lj}^{(q+1)}\geqslant a_{lk}a_{kj}^{(q)}$$

若$l\notin S$且$k\in S$，则$a_{lj}^{(q+1)}=0$且$a_{kj}^{(q)}>0$，由上面不等式推知$a_{lk}=0$。若记

$$S=\{l_1, l_2, \cdots, l_s\}$$

则可构造置换阵P如下：将单位阵I的第l_s列与其最后一列置换，

第l_{t-1}列与其倒数第二列置换，如此依次做下去，便得到P。这样便有

$$PAP^T=\begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

其中D为$s\times s$阶矩阵，且$1\leqslant s<k$。从而说明A是可约的，出现矛盾。

充分性。 假定对所有的脚偶(i,j)，存在某一自然数q使得$a_{ij}^{(q)}>0$，而$A\geqslant 0$却是可约的，将会导出矛盾。因为据A的可约性，存在置换阵P，使得

$$PAP^T=\begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$$

其中B与D是方阵。根据矩阵分块的乘法运算，对任一自然数q有

$$A^q=P^T\begin{bmatrix} B^q & 0 \\ C^* & D^q \end{bmatrix}P$$

因此在A^q中与右端右上角零块阵相对应的元为

$$a_{ij}^{(q)}=0,\ \forall q\in N$$

这与我们的假定相矛盾。

推论6.3.4 $n\times n$阶非负矩阵A是不可约的，其充分必要条件是存在一个自然数p，使得

$$A+A^2+\cdots+A^p>0$$

这个推论的证明留作练习。

如果注意到

$$a_{ij}^{(k)}=\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_{k-1}=1}^{n}a_{ij_1}a_{j_1j_2}\cdots a_{j_{k-1}j},\ a_{ij}\geqslant 0,$$ 则不难看出，$a_{ij}^{(k)}>0$的充分必要条件是存在一组$j_1,j_2,\cdots,j_{k-1}$，使得$a_{ij_1}a_{j_1j_2}\cdots a_{j_{k-1}j}>0$。于是便得到如下结论。

定理6.3.5 $n\times n$阶非负矩阵A是不可约的，其充分必要条件是

$$\sum_{k=1}^{n} A^k > 0$$

证明 充分性由推论6.3.4可得。现证必要性。若$A \geqslant 0$是不可约的，则据推论6.3.4，存在自然数p，使得

$$\sum_{k=1}^{p} A^k > 0$$

当$p \leqslant n$时，结论显然成立。若$p < n$，上式可写为

$$A\left(\sum_{k=0}^{p-1} A^k\right) > 0$$

其中$p-1 \geqslant n$。再利用Cayley-Hamilton定理，上式可写为

$$\sum_{k=1}^{n} c_k A^k > 0$$

欲使上述不等式成立，对每一给定的(i, j)，必存在某一自然数$q \leqslant n$，使得$(A^q)_{ij} > 0$，因此便证得

$$\sum_{k=1}^{n} A^k > 0$$

从这个定理的证明过程，可以看出定理6.3.3中的$q \leqslant n$。事实上，有如下推论。

推论6.3.6 设m为$A \geqslant 0$的最小多项式的次数，则定理6.3.3中的q满足

$$q \leqslant \begin{cases} m & i = j \\ m-1 & i \neq j \end{cases}$$

推论6.3.7 设m为不可约非负矩阵A的最小多项式的次数，则

$$(I+A)^{m-1} > 0$$

证明 把不等式左端展开

$$B = (I+A)^{m-1} = I + c_1 A^1 + \cdots + c_{m-2} A^{m-2} + A^{m-1}$$

其中$c_k(k=1, 2, \cdots, m-2)$皆大于零。利用推论6.3.6得

$$B_{ij}>0,\quad i\neq j$$

而$B_{ii}>0$是显然的。故$B>0$。

例1. 对于矩阵

$$A=\begin{pmatrix}0&0&0&1\\1&0&0&0\\0&0&1&1\\0&1&0&0\end{pmatrix}$$

由于$A+A^2+A^3+A^4\not>0$，因此A是可约的。

1912年德国数学家G·Frobenius把Perron定理推广到不可约的非负矩阵上。

定理6.3.8(Frobenius) 设$n\times n$阶矩阵$A\geqslant0$是不可约的。则

(1) 存在一个置换矩阵P及自然数$h\geqslant1$，使得

$$PAP^T=\begin{pmatrix}0&0&\cdots&0&A_h\\A_1&0&\cdots&0&0\\0&A_2&\cdots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&A_{h-1}&0\end{pmatrix}$$

其中对角线上的零块可以有不同的阶数。进一步还可得到

$$PA^hP^T=\mathrm{diag}(B_1,\ B_2,\ \cdots,\ B_h)$$

其中$B_i(i=1,\ 2,\ \cdots,\ h)$为非负矩阵。

(2) 矩阵$B_i(i=1,2,\cdots,h)$为素矩阵，且有相同的谱半径$\rho(A)^h$。

证明 这个定理的证明较长，且是构造性的，故分几步完成。首先定义集合

$$S=\{k\in N\mid a_{11}^{(k)}>0\}$$

并令h为S的最大公因子，即

$$h=\gcd(S)$$

集合S关于加法运算是封闭的，因为有

$$a_{11}^{(k+l)} \geqslant a_{11}^{(k)} a_{11}^{(l)}$$

于是h的充分大的倍数仍属于S，也就是说有

1° $a_{11}^{(kh)}>0$，对于充分大的$k\in N$。

事实上，还可证明下述结论：

2° 对每个$j\geqslant1$，存在整数$r_j(0\leqslant r_j<h)$，使得当$a_{j1}^{(l)}>0$时，对某个$k_0\in N$和充分大的k分别有

$$l=k_0h+r_j，\ a_{j1}^{(kh+r_j)}>0$$

首先固定j，定理6.3.3表明，存在t，$s\in N$使得

$$a_{j1}^{(t)}>0,\quad a_{1j}^{(s)}>0$$

设$t=hp+r_j$，由于对一切$k\in N$都有

$$a_{11}^{(k+s)} \geqslant a_{1j}^{(s)}\cdot a_{j1}^{(k)}$$

因此当$a_{j1}^{(k)}>0$时，$k+s$为h的整数倍，即

$$k+s\in S$$

同理，当$a_{j1}^{(l)}>0$时，$l+s$与$t+s$都是h的整数倍，故存在某个$q\geqslant0$使得

$$l=t+hq$$

又因$t=hp+r_j$，故有

$$l=h(p+q)+r_j=hk_0+r_j$$

反之，当$l=hk+r_j$且$k\geqslant p$时，便有

$$l=h(k-p)+t$$

但因 $$a_{j1}^{(l)} \geqslant a_{j1}^{(t)} a_{j1}^{(h(k-p))},\ a_{j1}^{(t)}>0,$$

且k足够大时$h(k-p)\in S$，从而有$a_{j1}^{(l)}>0$。

令$C_r=\bigcup\limits_{k\geqslant0}\{j|a_{j1}^{(hk+r)}>0\}$，$r=0，1，\cdots,h-1$,则所有的集合$C_r$都是非空的，否则$A'$的第1列，从而$A$的第1列都是零元素，这样$A$将是可约的了。从前面的证明可知,对所有的$j\geqslant1$，存在$r_j\in\{0，1，\cdots，h-1\}$及足够大的$k$使得

$$a_{j1}^{(hk+r_j)}>0$$

因而有

$$\bigcup_{r=0}^{h-1} C_r = \{1,\ 2,\ \cdots,\ n\}$$

若 $j \in C_r \cap C_s$，则 $a_{j1}^{(hq+r)} > 0$，$a_{j1}^{(hq'+s)} > 0$。由 2° 可知有

$$hq + r = hk_0 + r_j,\ hq' + s = hk_0' + r_j$$

据定义r_j，r，$s \in \{0,\ 1,\ \cdots,\ h-1\}$，因此必有

$$r_j = r = s$$

即诸集合 C_r（$r=0,\ 1,\ \cdots,\ h-1$）是两两不交的。一般情况下，有

3° 若$a_{ij}>0$且$j \in C_r$，则有

$$i \in \begin{cases} C_{r+1} & 0 \leqslant r < h-1 \\ C_0 & r = h-1 \end{cases}$$

这个结论之所以成立，是因为$a_{i1}^{(k+1)} \geqslant a_{ij} a_{j1}^{(k)}$

有了上述准备，我们来构造置换阵P。设C_r中共有c_r个元，并令$C_r = \{i_1^r,\ i_2^r,\ \cdots,\ i_{c_r}^r\}$。置换阵$P$的前$c_0$列取同阶单位阵$I$的第$i_1^0$，$i_2^0$，$\cdots$，$i_{c_0}^0$列，紧接下来的$c_1$列取同阶单位阵$I$的第$i_1^1$，$i_2^1$，$\cdots$，$i_{c_1}^1$列，如此下去，$P$的最后$C_{h-1}$列，取同阶单位阵$I$的第$i_1^{h-1}$，$i_2^{h-1}$，$\cdots$，$i_{c_{h-1}}^{h-1}$列。根据3°的结论，便知有

$$P^T A P = \begin{pmatrix} 0 & 0 & 0 & A_h \\ A_1 & 0 & \cdots 0 & 0 \\ 0 & A_2 & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots A_{h-1} & 0 \end{pmatrix}$$

其中对角线上为$c_r \times c_r$阶的零矩阵，$P^T A P$中所有的正元素都位于诸A_r($r=1,\ 2,\ \cdots,\ h$)之中。上式的一个等价形式为

$$P A P^T = \begin{pmatrix} 0 & A_1 & 0 & \cdots 0 \\ 0 & 0 & A_2 & \cdots 0 \\ \vdots & \vdots & \vdots & \ddots \vdots \\ \vdots & \vdots & \vdots & A_{h-1} \\ A_h & 0 & 0 & \cdots 0 \end{pmatrix}$$

有了上面的结论，可以直接验证有下述结果：

4° $PA^hP^T=\text{diag}(B_1, B_2, \cdots, B_h)$

其中$B_i=A_{i-1}A_{i-2}\cdots A_1A_hA_{h-1}\cdots A_i$，$i=1, \cdots, h$

下面证明定理的第二部分。假定 $j\in C_0$，则在2° 的证明中，只要行列互换就得到如下结论：对某个正整数k_j有$a_{1j}^{(hk_j)}>0$。从1°知，对一切足够大的自然数 k 有$a_{11}^{(hk)}>0$。因此，对这样的一些k也有

$$a_{1j}^{(h(k_j+k))}>0$$

故存在一个与j无关的k_0，使得对所有的$j\in C_0$有

$$a_{1j}^{(hk_0)}>0$$

类似地可证：当$i\in C_0$时，存在一个与i无关的 l_0，使得对所有的$i\in C_0$也都有

$$a_{i1}^{(hl_0)}>0$$

由于有

$$a_{ij}^{(h(k_0+l_0))}\geqslant a_{i1}^{(hl_0)}a_{1j}^{(hk_0)}$$

故有 $$a_{ij}^{(h(k_0+l_0))}>0 \qquad \forall i, j\in C_0$$

从而知$B_1=A_hA_{h-1}\cdots A_1$为素矩阵。据4°，对所有的k有

$$B_2^{k+1}=A_1B_1^kA_hA_{h-1}\cdots A_2$$

当$B_1^k>0$时，必有$B_2^{k+1}>0$。这表明B_2是素矩阵。类似可证B_3，…，B_h皆为素矩阵。由于B_1，B_2是素矩阵，故存在向量$p>0$，$q>0$，满足

$$B_1p=\rho(B_1)p, \quad B_2q=\rho(B_2)q$$

又因为有等式

$$A_1B_1=B_2A_1, \quad A_1B_1p=\rho(B_1)A_1p$$

故有 $$B_2(A_1p)=\rho(B_1)(A_1p)$$

这表明$\rho(B_1)$是B_2的相应于正特征向量A_1p 的特征值。由于B为素矩阵，故存在$k\in N$，使得

$$B_2^k>0$$

但据上述等式又有

$$B_2^k(A_1p)=\rho(B_1)^k(A_1p)$$

据正矩阵的Perron定理得（注意$\rho(B_2^h)=\rho(B_2)^h$）

$$\rho(B_1)=\rho(B_2)$$

类似可证$\rho(B_i)=\rho(B_1)$，$i=1, 2, \cdots, h$。据4°得

$$\rho(PA^hP^T)=\rho(A^h)=[\rho(A)]^h$$

$$=\rho(\text{diag}(B_1, B_2, \cdots, B_h))=\rho(B_1)$$

至此定理证完。

推论6.3.9 若$n\times n$阶非负矩阵A是不可约的，则$\rho(A)$为A的具有正特征向量x的单特征值。并且所有与$\rho(A)$有相同模的其它特征值也是单重的，它们是方程

$$\lambda^h-\rho(A)^h=0$$

的h个互异根，A的任一非负特征向量都为x的一个标量倍。

证明 这里只证明定理的前半部分，其余部分請参考Berman和Plemmons的著作。

从Frobenius定理可知，存在正向量y满足

$$PA^hP^Ty=\rho(A)^hy$$

于是有

$$A^h(P^Ty)=\rho(A)^h(P^Ty)$$

若取$x=P^Ty$，显然$x>0$，且

$$A^hx=\rho(A)^hx$$

令

$$\hat{x}=\sum_{i=0}^{h-1}\rho(A)^{-i}A^ix$$

则有

$$A\hat{x}=\sum_{i=0}^{h-1}\rho(A)^{-i}A^{i+1}x=$$

$$=\sum_{i=0}^{h-2}(A)^{-i}A^{i+1}x+\rho(A)^{1-h}A^hx$$

$$=\sum_{i=1}^{h-1}\rho(A)^{-i+1}A^ix+\rho(A)x$$

$$=\rho(A)\left(x+\sum_{i=1}^{h-1}\rho(A)^{-i}A^ix\right)=\rho(A)\hat{x}$$

由于z是正向量之和，故$z>0$，这表明$\rho(A)$是A的具有正特征向量z的特征值。另一方面，由于$\rho(A)^h$是矩阵$\mathrm{diag}(B_1, B_2, \cdots, B_h)$的单特征值（因为$B_i$是素矩阵），从而知$\rho(A)^h$也是$PA^hP^T$及$A^h$的单特征值。前面已经指出，$\rho(A)^h$作为$A^h$的特征根的重数不小于$\rho(A)$作为$A$的特征根的重数，从而证明了$\rho(A)$为$A$的单特征值。

推论6.3.9充分地显示出定理6.3.8的含义。

定义6.3.10 对于$n\times n$阶非负不可约矩阵A，它的模等于谱半径$\rho(A)$的特征值的个数，叫做A的循环指数。如果$h>1$，则称A是按指数h循环的。

当A是素矩阵时，有$h=1$。

现在我们考察不可约的非负矩阵A的谱半径。对任意的非负非零向量x，我们定义

$$r_x=\min_{x_i\neq 0}\frac{(Ax)_i}{x_i}$$

显然有$r_x\geqslant 0$，并且r_x为使$\rho x\leqslant Ax$成立的最大实数ρ，于是有

定理6.3.11 若令$r=r_z=\max\limits_{x>0}r_x$，则有

(1) $r>0$；

(2) $Az=rz$；

(3) $z>0$。

证明 (1) 取 $u=(1, 1, \cdots, 1)^T$，$r_u=\min\limits_i\sum\limits_{j=1}^n a_{ij}$。由于不可约矩阵的每一行都不能为零向量，故$r_u\neq 0$，从而有$r\geqslant r_u>0$。

(2) 若$Az\neq rz$，则有$0\neq Az-r_z\geqslant 0$。现取$x=(I+A)^{n-1}z$，则由于$(I+A)^{n-1}>0$，所以$x>0$。另一方面又有$(I+A)^{n-1}(Az-rz)>0$，亦即

$$Ax-rx>0$$

这与r的定义发生矛盾。

(3) 由于$Az=rz$,故 $0<(I+A)^{n-1}z=(1+r)^{n-1}z$,从而有$z>0$。

从推论6.3.9可知，不可约非负矩阵 A 的唯一的正特征向量是与$\rho(A)$对应的。由此便推得定理6.3.11中的 $r=\rho(A)$。于是便有

定理6.3.12 若A是不可约非负矩阵，则

$$\rho(A)=\max_{0\neq x\geqslant 0}\ \min_{x_i>0}\ \frac{(Ax)_i}{x_i}$$

类似地可证明下述结论：

定理6.3.13 若A是不可约非负矩阵，则

$$\rho(A)=\min_{0\neq x\geqslant 0}\ \max_{x_i>0}\ \frac{(Ax)_i}{x_i}$$

且当$\rho(A)=\max\limits_{z_i>0}\dfrac{(Az)_i}{z_i}$时，$z$为$A$的对应于 $\rho(A)$ 的一个特征向量。

若记A的伴随矩阵为adj$A$，则还可证明

推论6.3.14 若$A\geqslant 0$是不可约的，$B(\lambda)=\mathrm{adj}(\lambda I-A)$，则有$B(\rho(A))>0$。

证明 设A的特征多项式为$f(\lambda)$，则由于$\rho(A)$是$f(\lambda)$的单根，故有

$$f(\rho(A))=0,\ f'(\rho(A))>0$$

据恒等式$(\lambda I-A)B(\lambda)=f(\lambda)I$可知，$B(\rho(A))$的列向量都是$A$的对应于$\rho(A)$的特征向量，注意到$A$的特征根 $\rho(A)$是单根，故 $B(\rho(A))$ 的每个元素皆不为零，并且每列的元素有相同的符号。再作关于A^T讨论，可得出 $B(\rho(A))$的所有元素都同号。又由于$f'(\lambda)=\mathrm{tr}B(\lambda)$，故

$$\mathrm{tr}B(\rho(A))=f'(\rho(A))>0$$

这便证明了$B(\rho(A))>0$。

关于不可约矩阵$A\geqslant 0$的谱半径$\rho(A)$的界，有下述结论：

定理6.3.15 设 $A\geqslant 0$ 是不可约的，x为A的正特征向量，记$\gamma=\max\limits_{i,j}(x_i/x_j)$，$s_i=\sum\limits_{j=1}^{n}a_{ij}$，$s=\min\limits_i s_i$，$S=\max\limits_i s_i$，则有

$$s\leqslant\rho(A)\leqslant S，(S/s)^{1/2}\leqslant\gamma$$

上述不等式为等式的条件是$s=S$。

证明 将行列式

$$f(\lambda)=\begin{bmatrix}\lambda-a_{11} & -a_{12} & \cdots & -a_{1n}\\ -a_{21} & \lambda-a_{22} & \cdots & -a_{2n}\\ \vdots & \vdots & & \vdots\\ -a_{n1} & -a_{n2} & \cdots & \lambda-a_{nn}\end{bmatrix}$$

的前 $n-1$ 行加到最后一行，然后再按最后一行展开，并令 $\lambda=\rho(A)$代入，便得到

$$\sum_{j=1}^{n}(\rho(A)-s_j)B_{nj}(\rho(A))=0$$

从推论6.3.14知，$B_{nj}(\rho(A))>0$，上式便意味着

$$s\leqslant\rho(A)\leqslant S$$

同时也给出了等式成立的充要条件。若设

$$x_M=\max_i x_i，\quad x_m=\min_i x_i$$

则有 $$\rho(A)x_i=\sum_{j=1}^{n}a_{ij}x_j\geqslant\left(\sum_{j=1}^{n}a_{ij}\right)x_m=s_ix_m$$

因而对任意的i都有

$$\rho(A)/s_i\geqslant x_m/x_i\geqslant\frac{1}{\gamma}$$

类似地可证

$$\rho(A)x_i\leqslant s_ix_M，\quad\rho(A)/s_i\leqslant x_M/x_i\leqslant\gamma$$

特别在上面不等式中取$s_i=S$及$s_i=s$时，有

$$\rho(A)/S\geqslant 1/\gamma，\quad\rho(A)/s\leqslant\gamma$$

因此有

$$(S/s)^{1/2}\leqslant\gamma$$

欲使等式

$$\rho(A)x_i = \sum_{j=1}^{n} a_{ij}x_j = s_i x_m$$

对所有i成立，其充分必要条件是x的坐标分量相同。从而A的各行之和相等：$s_i = \rho(A)$。这时显然有$S/s = 1 = \gamma$。

在结束本节内容之前，我们举例来说明定理 6.3.8 中构造P阵的用意。

例2． 若考虑不可约阵

$$A = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

则可取置换阵

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

则有

$$PAP^T = \begin{pmatrix} 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{pmatrix} \qquad PA^2P^T = \begin{pmatrix} 4 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 4 & 2 \end{pmatrix}$$

于是定理6.3.8中的

$$B_1 = \begin{bmatrix} 4 & 2 \\ 1 & 1 \end{bmatrix}, \qquad B_2 = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix}$$

且有 $\rho(B_1) = \rho(B_2) = \dfrac{5+\sqrt{17}}{2} = \rho(A)^2$。

例3． 若$A \geqslant 0$是不可约的，且设$h = 3$，也就是说有置换阵P，使得

$$P^T AP=\begin{pmatrix} 0 & 0 & A_3 \\ A_1 & 0 & 0 \\ 0 & A_2 & 0 \end{pmatrix}$$

因而有

$$P^T A^2 P=\begin{pmatrix} B_1 & 0 & 0 \\ 0 & B_2 & 0 \\ 0 & 0 & B_3 \end{pmatrix}$$

其中 $B_1=A_3A_2A_1$， $B_2=A_1A_3A_2$， $B_3=A_2A_1A_3$，此处 B_i 为素矩阵。如果要估算 A^{1047}，可以作如下的变换：

$$A^{1047}=(A^3)^{349}=P\mathrm{diag}(B_1^{349},\ B_2^{349},\ B_3^{349})P^T$$

据定理6.2.4知，极限 $\lim\limits_{k\to\infty}(\rho(A)^{-3}B_i)^k$ 存在，记之为 L_i，则便有

$$L_2=\lim_{k\to\infty}(\rho(A)^{-3}B_2)^k=\rho(A)^{-3}A_1L_1A_3A_2$$

$$L_3=\lim_{k\to\infty}(\rho(A)^{-3}B_3)^k=\rho(A)^{-3}A_2A_1L_1A_3$$

这意味着，先据定理6.2.4求出 L_1，再据上面二式求出 L_2 及 L_3，从而可得 A^{1047} 的近似矩阵

$$\rho(A)^{1047}P\mathrm{diag}(L_1,\ L_2,\ L_3)P^T$$

例4. 考察不可约的非负矩阵

$$A=\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

欲求置换阵 P，使得 P^TAP 为标准形。若记 A^k 的第 1 列为 u_k，欲计算 u_k，只要顺序计算

$$u_2=Au_1,\quad u_3=Au_2,\quad \cdots$$

就行了。由于我们只对 u_k 中的非零元（记为＋号）感兴趣，所以

重复上述计算过程直到 u_k 的零元与非零元与 u_1 有相同的位置为止，例如

$$
\begin{array}{cccc}
u_1 & u_2 & u_3 & u_4 \\
\begin{pmatrix} 0 \\ + \\ + \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} &
\begin{pmatrix} 0 \\ 0 \\ 0 \\ + \\ 0 \\ + \\ + \end{pmatrix} &
\begin{pmatrix} + \\ 0 \\ 0 \\ 0 \\ + \\ 0 \\ 0 \end{pmatrix} &
\begin{pmatrix} 0 \\ + \\ + \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}
\end{array}
$$

这表明 $a_{j1}^{(4)}>0 \rightleftharpoons a_{j1}^{(1)}>0$；因此，对所有的 n 与 j，$a_{j1}^{(3n+1)}>0 \rightleftharpoons a_{j1}^{(1)}>0$；$a_{j1}^{(3n+2)}>0 \rightleftharpoons a_{j1}^{(2)}>0$。这便告诉我们 $h=3$，$C_1=\{2,3\}$，$C_2=\{4,6,7\}$，$C_0=\{1,5\}$。有了 C_r，就可以构造矩阵 P

$$
P=\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

不难算出

$$
P^TAP=\begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}=\begin{pmatrix}
0 & 0 & A_3 \\
A_1 & 0 & 0 \\
0 & A_2 & 0
\end{pmatrix}
$$

进一步还可算得

$$P^T A^3 P = \mathrm{diag}\left(\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right)$$

第四节　随机矩阵与Markov过程

本章的开始就给出了随机矩阵的例子，它是一类重要的非负矩阵。这一节我们给出随机矩阵的定义，并考察它的一些基本性质，介绍非负矩阵的各种应用。首先介绍齐次 Markov链。这样做的原因，一是随机过程的研究为非负矩阵的理论提供了最完美的应用；二是Markov 过程能很好地描述工程、商业、物理、生物等多方面的现象，有着广泛的应用。Markov链通常是作为描述具有许多固定状态的系统的一个理论模型，这个系统在单位时间间隔内按概率矩阵从某一状态转移到另一状态。

如果系统在时刻 $k-1(k\geqslant 1)$ 处于状态 s_i，则下一时刻将以概率$t_{ij}(k)$转移到状态s_j。转移概率的集合 $t_{ij}(k)$ 对所有的i，j及k都是已知的。一旦系统的初始状态已知，它就完全规定了系统的概率特性。如果系统的状态是有限的，则这个过程也是有限的。如果概率$t_{ij}(k)$对所有的$k\geqslant 1$，都与 k 无关，则这一过程称为是齐次的，这时随机过程的转移矩阵为$T=(t_{ij})$。

定义6.4.1　有限齐次 Markov 过程是指下列条件所构成的系统：

(1)　一个有限状态集合$S=\{s_1, s_2, \cdots, s_n\}$；

(2)　一个$n\times n$阶矩阵$T=(t_{ij})$，其中t_{ij}是系统从状态s_i转移到状态s_j的概率；

(3)　一个n维向量$\Pi^0=(\pi_1^0, \pi_2^0, \cdots, \pi_n^0)$，其中$\pi_i^0$是系统在初始时刻处于状态$s_i$的概率。

我们用记法$\mu=(T, \Pi^0)$表示这个有限齐次 Markov链。显然

Markov链的转移矩阵T应满足条件

$$t_{ij}\geqslant 0,\quad \sum_{j=1}^{n} t_{ij}=1,\qquad 1\leqslant i,\ j\leqslant n \tag{*}$$

定义6.4.2 一个$n\times n$阶矩阵$T=(t_{ij})$若满足(*)中的条件，便称之为（行）随机矩阵。如果还满足条件

$$\sum_{i=1}^{n} t_{ij}=1,\quad j=1,\ 2,\ \cdots,\ n$$

则称之为双随机矩阵。

为方便计，记向量$e=(1,\ 1,\ \cdots,\ 1)^T$。

定理6.4.3 随机矩阵的最大特征值为1。非负矩阵T为随机矩阵的充分必要条件是：e为T对应于特征值1的特征向量。

证明 由于任一非负矩阵都可看作正矩阵序列的极限，即若$A\geqslant 0$，则总有$A_m>0$使得

$$\lim_{m\to\infty} A_m=A$$

又由于正矩阵A_m是不可约的，因此关于不可约非负矩阵的某些性质很容易推广到一般的非负矩阵。所以定理6.3.15中的不等式

$$s\leqslant\rho(A)\leqslant S$$

对所有非负矩阵也是成立的。当然，用极限的方法所得到的结论一般总是比原先的结论弱些，因为极限过程不能再保持严格不等式的成立。由于随机矩阵的行和相等，因而推得T的谱半径为1，显然e是T的对应于$\rho(T)=1$的特征向量。反之，若$Te=e$，即说明T的行和等于1，而T又是非负矩阵，故是随机矩阵。

定理6.4.4 若$A\geqslant 0$，$\rho(A)>0$，$z>0$且满足

$$Az=\rho(A)z$$

则$\rho(A)^{-1}A$与一个随机矩阵相似。

证明 令$z=(z_1,\ z_2,\ \cdots,\ z_n)^T$，构造对角阵

$$D=\mathrm{diag}(z_1,\ z_2,\ \cdots,\ z_n)$$

则由于$z>0$，故D^{-1}存在。作矩阵

$$T=D^{-1}(\rho(A)^{-1}A)D$$

则由于$T\geqslant 0$且满足

$$Te=D^{-1}(\rho(A)^{-1}A)De=D^{-1}\rho(A)^{-1}Az$$
$$=D^{-1}z=e$$

故T是一个随机矩阵

定理6.4.5 若A为随机矩阵，$\lambda\neq 1$为A的一个特征值，则

$$|\lambda|\leqslant\min\left\{1-\sum_j\min_i a_{ij},\ \sum_j\max_i a_{ij}-1\right\}$$

证明 设v为A^T的对应于λ的特征向量，即满足

$$v^TA=\lambda v^T$$

由$Ae=e$可得$v^Te=v^TAe=\lambda v^Te$。由于$\lambda\neq 1$，故$v^Te=0$。这意味着v与e是正交的。记

$$c_j=\min_i a_{ij},\quad j=1,\ 2,\ \cdots,\ n$$

$$c=(c_1,\ c_2,\ \cdots,\ c_n)^T,\quad A_c=(a_{ij}-c_j)=A-ec^T$$

则$A_c\geqslant 0$。由于$v^Te=0$，故有

$$\lambda v^T=v^TA=v^T(A-ec^T)=v^TA_c$$

这意味着λ为A_c的一个特征值，于是有

$$|\lambda|\leqslant\rho(A_c)$$

而A_c的所有行和为

$$\sum_{j=1}^n(a_{ij}-c_j)=1-\sum_{j=1}^n c_j$$

这就证明了

$$|\lambda|\leqslant\rho(A_c)=1-\sum_{j=1}^n c_j=1-\sum_j\min_i a_{ij}$$

类似的方法可证明

$$|\lambda|\leqslant\rho(A_c)=\sum_j\max_i a_{ij}-1$$

从而证明了定理6.4.5。

让我们回到Markov链上来。由于Π^0在概率的意义上定义了系统的初始状态，所以又把它叫做初始概率分布向量。一般情况下，若π_i^k表示k步后系统处于状态s_i的概率，则把

$$\Pi^k=(\pi_1^k,\ \pi_2^k,\ \cdots,\ \pi_n^k)$$

叫做第k个概率分布向量。显然也应该满足

$$\pi_i^k\geqslant 0,\ \sum_{i=1}^{n}\pi_i^k=1,\ k=0,\ 1,\ 2,\ \cdots$$

注意，这里把Π^k视为行向量。有时会遇到Π^k的分量之和不等于1的情况，这时便称之为分布向量，而略去概率两字。

引理6.4.6 设$\mu=(T,\ \Pi^0)$是一个初态为Π^0，转移矩阵为T的Markov链，则对一切正整数k有

$$\Pi^k=\Pi^{k-1}T=\Pi^0T^k$$

式中Π^k是行向量，其第i个分量是k步后过程处于状态s_i的概率。

证明 设k为一正整数，由于过程在第k步处于状态s_i的概率完全由它前一步所处的状态s_m所决定，因而系统从$k-1$时刻的状态s_m转移至第k时刻的状态s_i的概率为$\pi_m^{k-1}\cdot t_{mi}$；而系统在k时刻处于状态s_i的总概率显然应为上述概率之和

$$\pi_i^k=\sum_{m=1}^{n}\pi_m^{k-1}t_{mi}$$

据此便得到引理的结论。

在随机过程的研究中，最感兴趣的问题之一是过程的平稳性，其定义为：

定义6.4.7 如果一个Markov链$\mu=(T,\ \Pi^0)$满足：

$$\Pi^k=\Pi^0,\ k=1,\ 2,\ \cdots$$

则称其初始（概率）分布向量Π^0是平稳的，或把Π^0叫做平稳的（概率）分布向量。

例1. 考虑一个假想的人口流动问题。若某一地区有专门人才1800人，分布在A，B，C三个单位。每年每个单位把所有人才各分$\frac{1}{2}$与其它两个单位交流。今年A，B，C三个单位各有人才分别为200人，600人，1000人，问明年的分布情况如何？后

年又怎样？很多年后各单位的人口期望值又是多少？若用a^k，b^k，c^k分别表示k年后A，B，C三个单位的人才，即分布向量为

$$H^k=(a^k,\ b^k,\ c^k)$$

这显然是一个具有三状态的Markov链，转移矩阵及初始向量分别为

$$T=\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix},\quad H^0=(200,\ 600,\ 1000)$$

因此有

$$H^1=(800,\ 600,\ 400),\quad H^2=(500,\ 600,\ 700)$$

用归纳法可以证明

$$T^k=\begin{pmatrix} t_k & t_{k+1} & t_{k+1} \\ t_{k+1} & t_k & t_{k+1} \\ t_{k+1} & t_{k+1} & t_k \end{pmatrix}\quad t_i=\frac{1}{3}\left[1+\frac{(-1)^i}{2^{i-1}}\right]$$

因此T是幂收敛的，且

$$\lim_{k\to\infty}T^k=\frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

由此可见

$$H=\lim_{k\to\infty}H^k=(600,\ 600,\ 600)$$

它是这个Markov链唯一的稳态分布向量，其对问题的解释是明显的。

例2. 考虑一个沿直线作单位步长运动的粒子。假定粒子右移一步的概率为p，左移一步的概率为q。并假定它只能取有限个状态$\{s_1,\ s_2,\ \cdots,\ s_n\}$，$s_1$与$s_n$为端点。当粒子移到端点

时，便以概率1返回（这种情况叫做带反射壁的马尔科夫链）。

上述过程可用一个Markov链来描述，其转移矩阵为

$$T=\begin{pmatrix} 0 & 1 & & & & \\ q & 0 & p & & & \\ & q & 0 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & q & 0 & p \\ & & & & 1 & 0 \end{pmatrix}$$

显然$0<p<1$，$q=1-p$。这种链称为随机游动，能描述很多物理现象。T是不可约的，后面将对它作进一步的讨论。

例3. 二人做纸牌游戏。每个盒中装有三张纸牌，其中两张标有“赢”字，一张标有“输”字。由A从盒中抽出一张，若是赢牌，B便付给A一元錢；否则，A付与B一元錢。开始之前，A有一元錢，而B有三元錢。并约定在任何一方输光时，游戏结束。们我希望求出他们每人成为赢者的概率；还希望求出需要第十个牌盒的概率，也就是在第九次抽出纸牌后，他们二人手中都还有錢的概率。

我们把A作为研究的对象，他手中的錢作为状态。便自然地列出所有的状态：

s_1—A无有錢；s_2—A有1元錢；

s_3—A有2元錢；s_4—A有3元錢；

s_5—A有4元錢，即两人全部的錢之和。

由此可得到状态转移矩阵为

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

这种不假思索的描述不一定是最方便的。若把状态的顺序改变为：

$$s_1=0,\ s_2=4,\ s_3=1,\ s_4=2,\ s_5=3$$

则转移矩阵为

$$T=\left(\begin{array}{cc:ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hdashline \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \end{array}\right)=\begin{bmatrix} I & 0 \\ B & C \end{bmatrix}$$

可见 T 是一个可约矩阵。象在例 1 中所做的一样，可得到

$$T^k=\begin{pmatrix} I & 0 \\ \sum_{i=0}^{k-1} C^i B & C^k \end{pmatrix}$$

此外，$\rho(C)<1$，故有

$$\lim_{k\to\infty} C^k=0$$

$$\lim_{k\to\infty}\sum_{i=0}^{k-1} C^i B=(I-C)^{-1}B=\frac{1}{15}\begin{pmatrix} 7 & 8 \\ 3 & 12 \\ 1 & 14 \end{pmatrix}$$

$$\lim_{k\to\infty} T^k=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{7}{15} & \frac{8}{15} & 0 & 0 & 0 \\ \frac{3}{15} & \frac{12}{15} & 0 & 0 & 0 \\ \frac{1}{15} & \frac{14}{15} & 0 & 0 & 0 \end{pmatrix}$$

由于开始时，A处于状态s_3，因而Markov链的初态为

$$\Pi^0=(0,\ 0,\ 1,\ 0,\ 0)$$

据引理6.4.6，有

$$\Pi=\lim_{k\to\infty}\Pi^k=\lim_{k\to\infty}\Pi^0T^k=\left(\frac{7}{15},\ \frac{8}{15},\ 0,\ 0,\ 0\right)$$

它是$\mu=(T,\ \Pi^0)$的稳态分布向量。由此可以看出，A最后成为赢家（处于状态s_2）的概率为$\frac{8}{15}$，而最后成为输家（处于状态s_1）的概率为$\frac{7}{15}$。若要使这种游戏无休止地进行下去，则A处于状态s_3，s_4，s_5的概率总和不能为零。而从稳态分布向量来看，这种游戏不会无休止地进行下去。由于

$$C^9=\begin{bmatrix} 0 & \left(\frac{2}{3}\right)^9 & 0 \\ * & 0 & * \\ 0 & * & 0 \end{bmatrix}$$

因而9次抽牌后，A不可能处于状态s_3与s_5，处于状态s_4的概率（也就是需要第10个牌盒的概率）为$\left(\frac{2}{3}\right)^9\approx0.026$。

通过上面的几个例子，对Markov链有了初步的了解。现在回到状态集合为$\{s_1,\ s_2,\ \cdots,\ s_n\}$，转移矩阵为$T$的一般Markov链上来，并对状态之间的关系引入下述概念：

定义6.4.8 如果所考虑的过程经过有限步之后可从状态s_i转移到状态s_j，则称从s_i到s_j有通道，记作$s_i\to s_j$。如果s_i到s_j有通道且s_j到s_i也有通道，则称s_i与s_j是互通的，记作$s_i\leftarrow\rightarrow s_j$。

状态之间的互通关系是等价关系，因而可将状态集合按这一等价进行分类。有了这些概念，便可引入：

定义6.4.9 我们把一个Markov链视为由互通关系划分的等价类。所谓从类α到类β有通道（记作$\alpha\to\beta$），指的是存在$s_i\in\alpha$，$s_j\in\beta$使得$s_i\to s_j$。如果某一类与其它的类都无通道，则

称之为终类（*final class*）。如果终类只包含单一的状态，则称该状态为吸收态（*absorbing state*）。

从上述定义可知：$s_i \to s_j$ 当且仅当存在自然数 k，使得 $(T^k)_{ij} > 0$；状态 s_i 是吸收态，其充分必要条件是 $t_{ii} = 1$。

若 T 是可约的，则一定存在置换阵 P 使得

$$PTP^T = \begin{bmatrix} T_{11} & 0 & \cdots & 0 \\ T_{21} & T_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ T_{s1} & T_{s2} & \cdots & T_{ss} \end{bmatrix}$$

其中对角块 T_{ii} 是方阵且是不可约的或是1×1的 0 。每个对角块 T_{ii} 对应于 Markov 链的一个类（为什么？）T_{ii} 对应的类是终类，其充分必要条件为 $T_{ij} = 0$，$j = 1, 2, \cdots, i-1$。从 PTP^T 的块三角形可以看出，每个Markov 链至少有一个终类。由于不可约矩阵的块三角形式只含有一块（根据定义），故它所对应的Markov链只有一个等价类。也就是说：一个Markov链只有一（终）类，其充分必要条件是 T 为不可约的。

定义6.4.10 对状态 s_i 来说，若存在 s_j 使得 $s_i \to s_j$，但 $s_j \not\to s_i$；也就是说 s_i 有到 s_j 的通道而沒有 s_j 到 s_i 的通道，则称 s_i 为瞬态（*transient state*）。否则，便称 s_i 为遍历态（*ergodic state*）。s_i 为遍历态的充分必要条件是：$s_i \to s_j$ 蘊含着 $s_j \to s_i$。

从定义6.4.10可知，Markov 链中的某一类，如果有一个状态是瞬态（遍历态），则该类中的所有状态都是瞬态（遍历态）。

定义6.4.11 由互通关系在Markov 链的状态集合上所导出的类 α，如果包含一个瞬态，则称为瞬态类。如果包含一个遍历态，便称之为遍历类。因而，α 是遍历类的充分必要条件为 α 是终类（請读者证明之）。

作为这些概念的一个说明，例 1 及例 2 中的每个状态都是遍历态，因为它们的转移矩阵都是不可约的，这两个Markov 链各自构成单一的遍历类。而例 3 的情况就不同了，s_1 与 s_2 是吸收态。也就是说，当 A 沒有錢或有了 4 元錢时，游戏便中止。因而

$\{s_1\}$，$\{s_2\}$是两个遍历类。s_3，s_4，s_5都是瞬态，集合$\{s_3$，s_4，$s_5\}$构成一个瞬态类。

例4. 考虑一个Markov链，其转移矩阵为

$$
T=\begin{pmatrix}
\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\
0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{12} & \frac{1}{3} & \frac{1}{4} & 0 & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2}
\end{pmatrix}
$$

置换2，4行与2，4列，使得

$$
PTP^T=\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{3} & \frac{1}{12} & 0 & 0 \\
0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2}
\end{pmatrix}
$$

可见状态s_1，s_2，s_5是遍历态，并且s_5是一吸收态。而s_3，s_4，s_6是瞬态。因而$\{s_1$，$s_2\}$，$\{s_5\}$是遍历类，$\{s_3$，$s_4\}$，$\{s_6\}$是瞬态类。注意，这里s_i是经过置换后的状态。

定义6.4.12 如果一个Markov 链是由单一的遍历类构成，则称之为遍历的Markov链。如果在固定的某一步k，Markov链的任一状态都能以一个大于零的概率转移到任意其它状态，则称之为正则链。如果一个链是遍历的且是非正则的，则称之为周期链。

显然，正则链是遍历的，而遍历链却未必正则。例如转移矩阵为

$$T=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

的链是遍历的，但它不是正则的，因而是周期的。

定义6.4.13 若一个链的每个遍历类都是由单个吸收状态构成的，则称之为吸收链。

下面我们要用前面介绍过的非负矩阵，对 Markov 链进行分类。首先有

定理6.4.14 设$\mu=(T, \Pi^0)$是Markov链，则

(1) μ是遍历的$\rightleftarrows T$为不可约矩阵

(2) μ是正则的$\rightleftarrows T$为素矩阵

(3) μ是周期的$\rightleftarrows T$是不可约的循环矩阵，即循环指数$h>1$。

证明(1) 在定义 6.4.10 之前的一段说明中曾得到结论：Markov 链只有一个终类的充分必要条件为 T 是不可约的。据定义6.4.12，即知(1) 成立。

(2) 由于一个链是正则的充要条件为：$T^k>0$。因此据素矩阵的定义知(2) 成立。

(3) 是定义6.4.12及定义6.3.10的直接结果。

据定理6.4.14，例 1 中的链是正则的，例 2 中的链是循环的。下面定理给出了Markov 链的平稳概率分布向量的存在性。

定理6.4.15 任一Markov 链都有一个平稳概率分布向量。

证明 设转移矩阵为T，则$\rho(T)=1$。据定理 6.2.9，存在一个行向量$0\neq x\geqslant 0$满足$xT=x$。对x作标准化后得

$$\Pi=\left(\sum_{i=1}^{n}x_i\right)^{-1}x$$

由于对所有的 k 都有

$$\Pi^k=\Pi T^k=\Pi$$

因此上述Π就是该Markov链的平稳概率分布向量。

下面我们用矩阵T的特征来刻划Markov链的基本类型。

定理6.4.16 随机矩阵T是某一遍历 Markov 链的转移矩阵，其充分必要条件是：

(1) T以 1 为其单特征值，

(2) 存在一个行向量$x>0$使得

$$xT=x$$

且上述x除了一个正倍数外是唯一的。

证明这个定理需要用到不可约矩阵的如下结论。

引理6.4.17 矩阵 $A\geqslant 0$ 是不可约矩阵的充分必要条件：$\rho(A)$是A的单特征值，且

$$\{x\mid x>0,\ Ax=\rho(A)x\}=\{x\mid x>0,\ A^Tx=\rho(A)x\}$$

这个引理的证明略。有兴趣的读者可参阅A.Berman等的著作（$P42$）。

定理6.4.16的证明。据引理6.4.17，T是不可约的，其充分必要条件是：1为T的单特征值，并且T具有对应于特征值 1 正的左特征向量 x。另据推论 6.3.9，T 的任何与 1 对应的左特征向量皆为x的一个倍数。而定理 6.4.14的(I) 则表明T的不可约是遍历链的充分必要条件。

下面的结论是定理6.4.15及定理6.4.16的直接推论。

推论6.4.18 Markov 链是遍历的，其充分必要条件为它有唯一的正的平稳概率分布向量。

定理6.4.16把遍历的Markov链与矩阵T的特征联系了起来。那么正则的 Markov 链，其转移矩阵T的特征又如何呢？为回答这个问题，先给出一个引理。

引理6.4.19 不可约的非负矩阵 A 为素矩阵的充分必要条件是

$$\delta(A)<\rho(A)$$

其中 $$\delta(A)=\max\{|\lambda|\,|\,\lambda\in\sigma(A),\ \lambda\neq\rho(A)\}$$

这里$\sigma(A)$表示A的特征值的集合。

证明 据Frobenius定理知，不可约矩阵$A\geqslant0$为素矩阵的充分必要条件是：A的循环指数$h=1$。也就是说，A的所有特征值中，其模等于$\rho(A)$的特征值只有$\rho(A)$，故$\delta(A)<\rho(A)$。

定理6.4.20 随机矩阵 T 为某一正则的 Markov 链的转移矩阵，其充分必要条件是定理6.4.16中的(1)，(2) 成立，再加上条件

(3) $\delta(T)<1$。

证明 由于正则链是遍历的，所以把定理 6.4.16 与引理6.4.19结合起来，便证得定理6.4.20。

如果要问周期Markov 链的转移矩阵有什么特征？则有结论：

定理6.4.21 随机矩阵T为某一Markov 链的转移矩阵，其充分必要条件是定理6.4.16中的(1)，(2) 成立，再加上条件

(4) 存在置换阵P，使得

$$PTP^T=\begin{pmatrix}0 & T_1 & 0 & \cdots & 0\\ 0 & 0 & T_2 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & T_{h-1}\\ T_h & 0 & 0 & \cdots & 0\end{pmatrix}$$

其中$h>1$，对角线上的块是方的。

证明 Markov链为周期的充分必要条件是：T是不可约的非素矩阵。所以据 Frobenius 定理有$h>1$，即在单位圆上有h个特征根。

定义6.4.22 若Markov 链是周期的，则把其转移矩阵T的循环指数h叫做该链的周期。

如果一个Markov链是周期的。那么该链所描述的某系统若初始状态处于与T_i相应的状态集合，则下一时刻将处于与T_{i+1}相应的状态集合，这里$i=1, 2, \cdots, h-1$。如果在某一时刻处于与T_h相应的状态集合，下一时刻将转移到与T_1相应的状态集合中去。因而上述过程周而复始的循环（請读者考虑PT^2P^T）。

任一个具有n个状态的Markov链，假若它共有r个遍历类：S_1，$S_2, \cdots, S_r$每个遍历类S_i有n_i个状态，则该链共有$m=n_1+n_2+\cdots+n_r$个遍历态，$t=n-m$个瞬态。我们又知道该链的转移矩阵T可置换成如下形式

$$PTP^T=\begin{bmatrix} T_{11} & & & \\ T_{21} & T_{22} & & \\ \vdots & \vdots & \ddots & \\ T_{s1} & T_{s2} & \cdots & T_{ss} \end{bmatrix}$$

据定义6.4.9及定义6.4.11知，与T_{ii}相应的类是终类的充分必要条件是：$T_{ij}=0$，$j=1, 2, \cdots, i-1$。而一个类是遍历类的充分必要条件是：该类为终类。因而前述转移矩阵总可置换成

$$\begin{pmatrix} D_1 & & & & \\ 0 & D_2 & & & \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \cdots D_r & & \\ B_1 & B_2 & \cdots B_r & C \end{pmatrix} \qquad *$$

其中D_i是相应于类S_i的$n_i \times n_i$阶不可约阵（$i=1, 2, \cdots, r$），C为t阶方阵，它相应的状态皆为瞬态。与转移矩阵T相似的上述形式叫做T的标准形。例 4 中给出的T矩阵就是一个标准形。

如果要问：转移矩阵T的标准形有什么应用？回答是定理6.4.24。为此，先给出一个引理。

引理6.4.23 若随机矩阵T具有标准形式（$*$），则有

$$\rho(C)<1$$

证明 设

$$D=\mathrm{diag}(D_1, D_2\cdots D_r)$$

于是有

$$T^k=\begin{bmatrix} D^k & 0 \\ B^{(k)} & C^k \end{bmatrix}$$

其中$B^{(k)}$是由$B^{(1)}=(B_1, B_2, \cdots, B_r)$，$C$及$D$所确定的矩阵。由于任意的瞬态必须在有限步（比如在$k_i$步）之内转移到某个遍历态，故$B^{(k_i)}$中的每一行至少有一个正元，从而$C^{k_i}$的最大行和$S<1$，据定理6.3.15，有$\rho(C)<1$。

定理6.4.24 一个随机矩阵能作为某个吸收Markov链的转移矩阵，其充分必要条件是它能被置换成如下形式

$$\begin{bmatrix} I & 0 \\ B & C \end{bmatrix}$$

且 $\rho(C)<1$。

证明 据吸收的Markov链的定义，每个遍历类都是由单个的吸收态所构成。因式(*)中的$D_i=(1)$。利用引理6.4.23，便得定理的结论。

例3. 中的链就是一个吸收链，所给出的T就具有定理6.4.24所示的标准形式。

前面我们考虑了一个Markov链的分类与它的转移矩阵T之间的关系。那么它的平稳概率分布向量又是怎样的呢？有下述定理给出回答。

定理6.4.25 一个Markov链的每个遍历类S_i，相应地都存在一个唯一的平稳概率分布向量Π_i，并且Π_i具有如下性质：行向量Π_i在与S_i的状态相对应的位置上有正元素，而其它的元素皆为零。此外，该链的所有平稳概率分布向量Π都可表示为诸平稳概率分布向量Π_i的一个凸组合，即

$$\Pi=\sum_{i=1}^{r}\lambda_i\Pi_i, \quad \lambda_i\geqslant 0, \quad \sum_{i=1}^{r}\lambda_i=1$$

其中r是该链所具有的遍历类的个数。

证明 设T为所考虑的Markov链的转移矩阵。不失一般性，

假定它具有(*)所示的标准形。又设 $\theta(i)$ 是该链关于状态转移矩阵D_i 唯一的正平稳概率分布向量，这里 $i=1, 2, \cdots, r$，据推论6.4.18，这样的向量是存在且唯一的。于是下列r个n维向量

$$\Pi_1=(\theta(1), 0, 0, \cdots, 0)$$
$$\Pi_2=(0, \theta(2), 0, \cdots, 0)$$
$$\vdots$$
$$\Pi_r=(0, 0, 0, \cdots, \theta(r), 0)$$

是线性无关的平稳概率分布向量。显然它们的任一凸组合仍是平稳概率分布向量。留下来的是证明，该Markov链的任一平稳概率分布向量

$$\Pi=(z_1, z_2, \cdots, z_r, z_{r+1})$$

都能表示成 $\Pi_i(i=1, 2, \cdots r)$的一个凸组合。上述向量Π的分块是与T的标准形一致的。首先可知

$$z_{r+1}C=z_{r+1}$$

引理6.4.23告诉我们：$\rho(C)<1$，故 1 不是C的特征值，因此只有$z_{r+1}=0$。而对任意的$i=1, 2, \cdots, r$，又有

$$z_iD_i=z_i$$

再据定理6.4.16，又有

$$z_i=\alpha_i\theta(i), \quad i=1, 2, \cdots, r, \quad \alpha_i>0,$$

由于行向量Π及$\theta(i)(i=1, 2, \cdots, r)$的行元素之和等于 1，故有

$$\sum_{i=1}^{r}\alpha_i=1$$

从而证明了Π是诸Π_i的一个凸组合。

非负矩阵的理论对研究 M-矩阵有很大的帮助。但我们这里介绍的只是一个初步知识，对真正地去刻划M-矩阵还是不够的。限于篇幅，只好把M-矩阵及其在系统工程上的应用略去了。

第五节 习 题

1. 设$0 \neq x \geqslant 0$，$A > 0$，证明$Ax > 0$。

2. 给出一个$0 \neq A \geqslant 0$及$0 \neq x \geqslant 0$使得$Ax = 0$。

3. 设$A > 0$，$z \geqslant w$，证明 $Az \geqslant Aw$。

4. 若对一切$x \geqslant 0$都有$Ax \geqslant 0$，试证$A \geqslant 0$。

5. 若$A \geqslant B \geqslant 0$，试证$\rho(A) \geqslant \rho(B)$。

6. 用 2×2 的矩阵说明 $\rho(AB) \leqslant \rho(A)\rho(B)$ 不是总能成立的。

7. 设

$$A = \begin{pmatrix} 7 & 2 & 2 \\ 2 & 1 & 1 \\ 4 & 2 & 2 \end{pmatrix}$$

求$\rho(A)$及$\lim\limits_{n \to \infty}(\rho(A)^{-1}A)^n$。

8. 设$A \in \mathbf{R}^{n \times n}$（不必是非负矩阵），$\lambda > 0$为$A$的一个单特征值，且$A$的其它$n-1$个征特值的模都小于$\lambda$。证明极限$\lim\limits_{n \to \infty}(\lambda^{-1}A)^n$存在。若$y$为$A$的对应于$\lambda$的任一特征向量，$x$为$A^T$的对应于$\lambda$的任一特征向量。证明

$$\lim_{n \to \infty}(\lambda^{-1}A)^n = (x^T y)^{-1} y x^T$$

9. 给定矩阵

$$A = \begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$$

又知它的特征多项式为$(\lambda+2)^2(\lambda-4)$，利用习题8计算$\lim\limits_{n \to \infty}\left(\frac{1}{4}A\right)^n$。

10. 设$A \in \mathbf{C}^{n \times n}$，$\lambda$为$A$的$k$重特征根，$m \geqslant 1$。证明：

(1) λ^m是A^m的特征值。

(2) λ^m作为A^m的特征值，其重数不小于k。

(3) $\rho(A^m)=[\rho(A)]^m$

11． 若A为素矩阵，λ为A的任一具有正特征向量的特征值，证明$\lambda=\rho(A)$。

12． 设

$$0\leqslant A\leqslant\begin{bmatrix} 0_s & B \\ 0 & 0_t \end{bmatrix},\quad 0_s\in\mathbf{R}^{s\times s},\ 0_t\in\mathbf{R}^{t\times t}$$

且BC与CB均为正矩阵。试证：

(1) $\rho(BC)=\rho(CB)=\rho(A)^2$

(2) $(\rho(A)^{-1}A)^2$是幂收敛的，且极限$\lim\limits_{k\to\infty}(\rho(A)^{-1}A)^{2k}$的秩为2。

试求极限 $\lim\limits_{k\to\infty}(\rho(A)^{-1}A)^{2k+1}$。

13 若A是不可约的非负矩阵，且$\alpha_j\geqslant 0$。试证$\sum\limits_{j=1}^{k}\alpha_j A^j$也是不可约的。

14． 证明所有素矩阵都是不可约的。

15． 若$A\geqslant 0$是可约的。试证存在置换矩阵P，使得

$$PAP^T=\begin{bmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \vdots & \vdots & \ddots & \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix}$$

其中A_{ii}为方阵，且是不可约的，或为1×1阶的零矩阵。

16． 若A是不可约的。试证

$$\rho(A)=\max_{x>0}\min_{y>0}\frac{y^TAx}{y^Tx}=\min_{y>0}\max_{x>0}\frac{y^TAx}{y^Tx}$$

17．若$x>0$为不可约矩阵$B\geqslant 0$的特征向量，$A\geqslant 0$与B可换，即$AB=BA$。试证x也是A的一个特征向量。

18． 若$A\geqslant 0$，试证

$$\rho\left(\frac{A+A^T}{2}\right)\geqslant\rho(A)$$

其中等式成立的充分必要条件是：A与A^T关于特征值 $\rho(A)$有同一个特征向量。

19． 若 $A\geqslant 0$ 是不可约的。证明：当 $\lambda>\rho(A)$ 时有 $(\lambda I-A)^{-1}>0$。

20． 计算下列矩阵

$$A=\begin{pmatrix} 0 & B & 0 & 0 \\ 0 & 0 & B & 0 \\ 0 & 0 & 0 & B \\ C & 0 & 0 & 0 \end{pmatrix}$$

的特征值，其中

$$B=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad C=\begin{bmatrix} 6 & 13 \\ 4 & 6 \end{bmatrix}$$

21． 将下列矩阵

$$A=\begin{pmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix}$$

化成习题15中的形式，并求出谱半径$\rho(A)$。

22． 若$A\geqslant 0$为不可约矩阵。证明：极限

$$\lim_{n\to\infty}(\rho(A)^{-1}A)^n$$

存在的充分必要条件是A为素矩阵。

23． 若P是一个随机矩阵。证明：极限

$$\lim_{n\to\infty}P^n$$

存在的充分必要条件是：P 的特征值满足 $|\lambda|=1$意味着$\lambda=1$。

24． 若$A\geqslant 0$是不可约矩阵。证明：等式

$$\lim_{m\to\infty}[(A^m)_{ij}]^{1/m}=\rho(A)$$

成立的充要条件是A为素矩阵。

25．设$A\geqslant 0$是$n\times n$阶的素矩阵，$x^{(0)}$为任一正向量，定义

$$x^{(r)}=Ax^{(r-1)}=A^r x^{(0)}$$

$$\bar{\lambda}_r=\max_{1<i<n}\frac{x_i^{(r+1)}}{x_i^{(r)}},\quad \underline{\lambda}_r=\min_{1<i<n}\frac{x_i^{(r+1)}}{x_i^{(r)}}$$

试证

(1) $\underline{\lambda}_0\leqslant\underline{\lambda}_1\leqslant\cdots\leqslant\underline{\lambda}_r\leqslant\cdots\leqslant\rho(A)\leqslant\cdots\leqslant\bar{\lambda}_r\leqslant\cdots\leqslant\bar{\lambda}_0$；

(2) 序列$\{\bar{\lambda}_r\}$与$\{\underline{\lambda}_r\}$对任意的$x^{(0)}$都收敛于$\rho(A)$的充分必要条件为：$A$是素矩阵。

26．对矩阵A来说，若不存在置换阵P与Q使得

$$PAQ=\begin{bmatrix} B & 0\\ C & D\end{bmatrix}$$

则称A为完全不可约的，其中B与D为方阵。试证：

(1) A为完全不可约的分充必要条件是：对某个置换阵P，PA是不可约的，且主对角线上有非零元。

(2) 若$A\geqslant 0$是完全不可约的，则A是素矩阵，但相反的结论不必成立。

27．给出两个可约矩阵其积为正矩阵；再给出两个素矩阵其积为可约矩阵。并证明非负完全不可约矩阵之积仍是完全不可约矩阵。

28．指出下列矩阵中哪些是完全不可约的？哪些是素矩阵？哪些是循环矩阵？循环指数是多少？哪些是可约矩阵？

$$A_1=\begin{bmatrix} 0 & 1\\ 1 & 1\end{bmatrix},\quad A_2=\begin{bmatrix} 1 & 0\\ 1 & 1\end{bmatrix},\quad A_3=\begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix}$$

$$A_4=\begin{pmatrix} 0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{pmatrix},\qquad A_5=\begin{pmatrix} 0 & 1 & 0\\ 1 & 1 & 1\\ 0 & 1 & 0\end{pmatrix}$$

$$A_6=\begin{pmatrix}0&1&0&0\\0&0&0&1\\0&0&1&0\\1&0&0&0\end{pmatrix},\quad A_7=\begin{pmatrix}1&1&0&0\\0&1&1&0\\0&0&1&1\\1&0&0&1\end{pmatrix}$$

$$A_8=\begin{pmatrix}1&0&0&1\\0&1&1&0\\0&1&1&0\\1&0&0&1\end{pmatrix},\quad A_9=\begin{pmatrix}0&0&0&1\\0&0&1&0\\1&1&0&0\\1&0&0&0\end{pmatrix}$$

$$A_{10}=\begin{pmatrix}0&0&1&0\\0&0&1&1\\1&1&0&0\\0&1&0&0\end{pmatrix}$$

29． 若$A\geqslant 0$的关于谱半径$\rho(A)$的特征向量是正的。试证：A的所有满足$|\lambda|=\rho(A)$的特征值λ对应的初等因子都是一次的。

提示：先就随机矩阵证明这一结论。

30． 若P，Q都是随机矩阵。证明：对任意的$\lambda\in[0,1]$，$\lambda P+(1-\lambda)Q$也是随机矩阵。从而说明随机矩阵的集合是$\mathbf{R}^{n\times n}$中的凸集。

31． 考虑一个抛硬币的过程：每一次抛起两枚硬币，s_1代表同时出现正面，s_2代表出现一正一反，s_3代表同时出现反面。

(1) 证明这是一个Markov过程，并给出转移矩阵T。

(2) 如果在某次抛掷中出现了两个正面，试问三次之后出现两个反面的概率是多少？

(3) 对该Markov链的状态作一下分类，即指出哪些状态是遍历态，哪些状态是瞬态。

32． 若T为一Markov链的非奇异的转移矩阵，且$T^{-1}\geqslant 0$。试证T^{-1}也是某一Markov链的转移矩阵。

33． 若T与T^{-1}都是 Markov 链的转移矩阵。试证T是一个置

换阵。

34． 两个Markov链的转移矩阵分别为

$$T_1=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$T_2=\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

(1) 确定每个链的遍历态及瞬态。

(2) 确定每个转移矩阵的如下标准形：

$$\begin{pmatrix} D_1 & & & & \\ 0 & D_2 & & & \\ 0 & 0 & \ddots & & \\ \vdots & \vdots & \cdots & D_r & \\ B_1 & B_2 & \cdots & B_r & C \end{pmatrix}$$

35． 若T为一Markov链的转移矩阵，又设

$$\bar{T}=\frac{1}{2}(I+T)$$

试证：T与$\bar{T}$有相同的遍历态及瞬态。

36． 考虑有下面转移矩阵的Markov链：

$$T=\begin{pmatrix}\frac{1}{2} & 0 & \frac{1}{2}\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}$$

(1) 证明该链是正则的。

(2) 利用$\Pi T=\Pi$及$\pi_1+\pi_2+\pi_3=1$，求出该链的平稳概率分布向量。

(3) 求出$L=\lim\limits_{k\to\infty}T^k$。

37. 指出以下列矩阵为转移矩阵的Markov链中，哪些是遍历链？哪些是正则链？哪些是周期链？哪些是吸收链？

$$A_1=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix},\quad A_2=\begin{pmatrix}1 & 0\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

$$A_3=\begin{pmatrix}0 & \frac{1}{2} & \frac{1}{2}\\ 1 & 0 & 0\\ 1 & 0 & 0\end{pmatrix},\quad A_4=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ \frac{1}{2} & \frac{1}{2} & 0\end{pmatrix}$$

$$A_5=\begin{pmatrix}0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3}\\ 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4}\\ 0 & 0 & 1 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

38. 将习题37中的转移矩阵化成标准形。

39. 证明：每一个概率分布向量Π都能找到一个正则的

arkov链，以Π为其平稳概率分布向量。

40． 一个Markov链的转移矩阵为T，并令

$$T^m=(t_{ij}^{(m)})$$

证明下述结论成立，或不成立时给出反例。

(1) 若s_i是瞬态且对某一个$m\in N$有$t_{ij}^{(m)}>0$，则s_j是一个瞬态。

(2) 若s_j是瞬态且对某一个$m\in N$有$t_{ij}^{(m)}>0$，则s_i是瞬态。

(3) 若T_1与T_2都是正则链的转移矩阵，则T_1T_2也是一正则链的转移矩阵。

(4) T_1与T_2同(3)所设，$\lambda\in(0,1)$矩阵

$$T=\lambda T_1+(1-\lambda)T_2$$

也为一正则链的转移矩阵。

41． 若具有n个状态的正则Markov链的转移矩阵T是双随机矩阵，则该链的平稳概率分布向量的每个元素都为$\frac{1}{n}$。

42． 若具有n个状态的Markov链的转移矩阵为T。试证：

(1) 该链是遍历链的充要条件是：以

$$\tilde{T}=\frac{1}{2}(I+T)$$

为转移矩阵的链是正则的。

(2) 这时$\tilde{T}^n>0$。

参 考 文 献

Hoffman, K., Kunze, R., Linear Algebra, Prentice Hall, New Jersey, 1961.

Gantmacher, F.R., Applications of the Theory of Matrices, vols. Ⅰ, Ⅱ, Chelsea, New York, 1959.

Bellman, R.: An Introduction to Matrix Analysis, 2nd ed., McGraw-Hill Book Co., New York, 1970.

Marcus, M., Minc, H.: A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Inc., Boston, 1964.

Chung, K.L.: Markov Chains with Stationary Transition Probabilities, Springer-Verlag, Berlin, 1960.

Rao, C.R., Mitra, S.K.: Generalized Inverse of Matrices and Its Applications, Wiley, 1971.

Ben-Israel, A., Greville, N.E.: Generalized Inverses: Theory and Applications, John Wiley&Sons, Inc. New Yo·k, 1974.

Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.

黄琳：系统与控制理论中的线性代数，科学出版社，1984.

屠伯埙：《矩阵的秩的下界与方阵的非异性(Ⅰ)、(Ⅱ)》，复旦大学学报，21，1982；23.1984.